Exploring vulnerability to infectious disease in a small-holder farming community in rural western Kenya.

William A. de Glanville

A thesis submitted for the degree of Doctor of Philosophy

University of Edinburgh

2014

Abstract

More than 2 billion people live on less than 2 US dollars per day. People in these conditions often have inadequate access to basic sanitation, safe water, and medical services. These individuals, households and communities may be at high risk for a wide range of preventable and treatable infectious diseases.

The aims of this study were to: 1) describe the prevalence of endemic helminth, protozoal, bacterial and viral infections of people in a small-holder farming community in western Kenya; 2) explore the spatial distribution of infection risk; 3) quantify associations between social and environmental conditions and individual- and household-level infection; 4) identify shared risk factors operating on multiple pathogens.

All data were collected between July 2010 and July 2012 as part of a cross-sectional survey of 416 households and 2113 people. This sample was considered representative of a population of 1.4 million people living in an area of western Kenya characterised by high levels of poverty. Sampled individuals were tested for exposure to, or infection with, 21 infectious agents using a range of faecal, blood and serological tests. Extensive questionnaire-based data were also collected.

Individual- and household-level risk factors for infection with prevalent pathogens were explored using multilevel logistic regression, with a particular focus on examining the impact of socioeconomic position (SEP). Hierarchical zero-inflated binomial (ZIB) regression was used to derive an estimate of household pathogen 'species richness' with correction for imperfect detection. This modelling framework allowed assessment of the relationship between household-level infection with each parasite and a range of social and environmental conditions and, uniquely for a single study setting, the average response of the 'group' of parasites to these conditions.

This study found very high levels of parasitism in the community, particularly with hookworm (36.3% (95% CI 32.8 – 39.9)), *Entamoeba histolytica/dispar* (30.1% (27.5 – 32.8)), *Plasmodium falciparum* (29.4% (26.8 – 32.0)), and *Taenia* spp. (19.7% (16.7 – 22.7)). Some degree of within-household clustering was found for all pathogens, and this was particularly large for the helminth species and HIV. Most pathogens also showed spatial heterogeneity in infection risk, with evidence of spatial clustering in household-level infection, most notably for HIV, *Schistosoma mansoni*, *P. falciparum* and the soil-transmitted helminths.

A socioeconomic gradient was identified, even in this predominantly poor community. Increasing socioeconomic position (SEP) resulted in significantly reduced risk of individual infection for *E. histolytica*/*dispar*, *P. falciparum*, and hookworm. By contrast, individuals living in the richest households were at significantly elevated risk of infection with *Mycobacterium* spp.. Individuals living in the poorest households were least likely to report the recent use of medical treatments.

The average pathogen species richness (out of 21 species) per household was 4.7 (range: 0 to 13). Following correction for detection error, the predicted average helminth species count (out of 6 species) was 3 (range: 0.94 to 5.96). While socioeconomic position had little effect on the probability that a household was infected with any of the helminth species of interest, domestic (within-household) transmission appeared to be greatest in the poorest households for hookworm, *S. mansoni*, *Ascaris lumbricoides* and *Strongyloides stercoralis*. Household size had a consistent effect on probably of household infection with each helminth species, so that the largest households were also the most pathogen diverse. Household-level helminth species richness was identified as a significant positive predictor of individual risk of HIV infection, raising potentially important questions about helminth-HIV interactions in the study area.

This study integrates approaches from epidemiology and ecology to explore infectious disease risk and its determinants at a range of social and geographic scales in a small-holder farming community in western Kenya. Considering risk at both the individual and household level within the same community can contribute to better understanding of the factors that influence disease transmission in both domestic and public domains.

Declaration of Authorship

I, William A. de Glanville, declare that the work presented within this thesis is my own. This work has not been submitted for any other degree or professional qualification.

Signed: .....

Date:.....

Acknowledgements

This thesis would not have been possible without the help and support of many, many people. Particular thanks to my supervisors, Eric Fèvre and Mark Bronsvoort. To Eric, for the opportunity to be involved in the kind of project I spent most of my undergraduate career staring out of the window and dreaming about, and particularly for allowing me to develop and explore such a wide range of ideas, some with more success than others! And to Mark B, upon whom I imposed myself for nearly a year and a half, and demanded much more time than he was probably expecting. His input has had a major role in shaping this thesis.

My continuing thanks to Kim Stevens and Dirk Pfeiffer, who set me on this path. To Ian Handel, Marc Kery, Cole Burton, and Gurutzeta Guillera-Arroita for discussions on zero inflated binomial models and how to code them, and to Nicola Wardrop and Paul Bessell for useful discussions on spatial approaches. I have been lucky to span three research groups in the UK and Kenya, and thanks for lots of interesting discussion and help along the way go to many, but particularly to Cheryl Gibbons, Claire Okell, Lian Thomas, Laure Made, Stella Mazeri, Sema Nickbakhsh, Amy Jennings and Jess Flood. And particularly to Annie Cook, my running mate in all of this.

This work was supported by a BBSRC Doctoral Training Grant, and the PAZ study on which the analysis is based was funded by the Wellcome Trust (085308). Thanks go to Sam Kariuki, the director of CMR KEMRI, for all he did to facilitate this work.

Final, and particular thanks to the PAZ team in Kenya: James Akoko, Lazarus Omoto, Lorren Alumasa, Daniel Cheriyot, Jenipher Ambaka, Fred Opinya, John Mwaniki, Hannah Kariuki, Gideon Mwali, George Omondi, Alice Kiyonga, Velma Kivali, Eunice Wauye, Lilian Abonyo, Maseno Cleophas, Charles Njau, and Bartholomew Wabwire, from all of whom I have learnt so much.

Table of Contents

List of Figures

List of Tables

List of abbreviations

AHP	Analytic hierarchy procedure
AIC	Akaike information criterion
BDFree	Bayesian disease freedom (software)
FEC	Formal ether concentration technique
ICC	Intracluster correlation co-efficient
KDE	Kernel density estimation
KK	Kato-Katz technique
LST	Land surface temperature
LTBI	Latent TB infection
MCA	Multiple correspondence analysis
MDA	Mass drug administration
MFA	Multiple factor analysis
MOR	Median odds ratio
NDVI	Normalised difference vegetation index
NTD	Neglected tropical disease
PAZ	The epidemiology of zoonoses in livestock and their keepers ("People, Animals and their Zoonoses") project
PCA	Principal component analysis
PCV	Proportional change in variance
RCS	Restricted cubic spline
RSA	Residual spatial autocorrelation
SDH	Social determinants of health
SEP	Socioeconomic position
STH	Soil transmitted helminth
VPC	Variance partition co-efficient
WHO	World Health Organisation
ZIB	Zero inflated binomial

Chapter 1

Introduction

> *"One can think of the middle of the twentieth century as the end of one of the most important social revolutions in history, the virtual elimination of the infectious diseases as a significant factor in social life."*
>
> Sir Frank Burnet. *Natural History of Infectious Disease*, 1962

1.1. The persistent burden of infectious disease: why research is still needed.

During 2010, there were 493,242 registered deaths in England and Wales (ONS 2014). Of these, 42,425, or 8.6%, could be attributed directly to an infectious cause.[1] In Kenya, by contrast, between 1994 and 2010, infectious disease was the cause of more than 64% of all mortality, with HIV/AIDS accounting for 29.3% of deaths; lower respiratory tract infections for 8.1%; TB for 6.3%; diarrhoeal diseases for 6.0%; and malaria for 5.8% (KHSSP 2012). In western Kenya, the region in which the work for this thesis was conducted, the child mortality rate (live born children not reaching their 5th birthday) has been estimated to be as high as 224 per 1000, and average life expectancy at birth just 38 years (Adazu et al. 2005). Preventable infectious disease, in combination with the effects of malnutrition, explains much of this excess mortality (Black et al. 2010). The stark gradient in infectious burden between developing and developed countries is a global phenomenon (Black et al. 2010; Liu et al. 2012), and one that means we are still a very long way from having "eliminated" infectious disease, which still afflicts billions and kills millions each year (Lopez & Mathers 2006).

Social inequality, weak infrastructure and material deprivation are likely to be at the root of much of the persistence of infectious disease as a major public health problem in poor countries (Alsan et al. 2011; Farmer 2001), as well as in some poor communities in rich countries (Brudney & Dobkin 1991; Hayward & Coker 2000; Hotez 2013). Whilst technological advancement has allowed most people, including many in the developing world (McMichael & Beaglehole 2000), to enjoy unprecedented increases in standards of

[1] ICD-10 codes used to extract counts of infectious causes of death from ONS (2014) data: A00-B99; G00; G04; G06; I00-I02; I30.1; 133.0; I40.0; J00-J06; J09-J18; J20-J22; L00-L08; M00-M03; M60.0; M86; N39.0; P35-P39.

sanitation, access to safe food and water, and to preventive and curative medical care, this is generally not the case for the world's 2.6 billion poorest people (WHO 2006a). Nutritional inadequacies, contaminated water, crowded housing and inadequate access to health services within impoverished populations can create the ideal conditions for endemicity for a very wide range of infectious diseases. Individual and structural poverty allows the 'old scourge' of TB and parasitic infections to persist (Farmer 2013; Hotez et al. 2007), emerging diseases to cause major outbreaks (IRIN 2012; IRIN 2014), or, as in the case of HIV, to become established and extremely widespread (Farmer 1996; Heymann 2005). Moreover, the impacts of infectious disease on health, productivity and education can contribute to Malthusian poverty traps (Ngonghala et al. 2014), condemning individuals, families and communities to continuous cycles of poverty. In economies that rely to a large extent on physical labour, endemic infectious disease is also likely to have had profound effects on national and regional development (Sachs 2011; Hotez et al. 2009).

The inequity of this situation, and the recognition of the responsibility of the global community for the most vulnerable and least served of its members, lead to the setting of the Millennium Development Goals (MDG) in 2000 (http://www.un.org/millenniumgoals/). Whilst MDG targets on poverty reduction by 2015 are unlikely to be met in sub-Saharan Africa, where 49% of people still live on less than 1.25 USD per day (World Bank 2014), massive international investment has seen some success towards MDG6, to "combat HIV/AIDS, malaria and other infectious diseases". The prevalence of HIV is now in decline in the region, having gone from 5.9% in 2001 to 4.9% in 2011 (UNDP 2013). Cases of malaria and TB, still a vast problem and a major cause of mortality (Murray et al. 2012; WHO 2013b), have also been reduced (UNDP 2013). Whilst this is a significant public health achievement, and one that requires substantial on-going international effort, the focus on these 'big 3' infectious diseases, and relegation of all else to the 'other' category, was considered by some to be representative of the on-going neglect of a group of pathogens that are the most common infections in poor populations (Molyneux et al. 2005; Liese & Schubert 2009; Molyneux 2008). This group, subsequently known as the 'neglected tropical diseases' (NTDs), contains a range of viral, bacterial, helminth and protozoal infections that have diverse life cycles and transmission routes, but which predominantly afflict individuals living in poverty and have hitherto been relatively neglected by research and policy makers. Although the NTDs tend not to be major causes of mortality (which in part explains their neglect), some estimates put their collective impact in terms of disability adjusted life years (DALYs) as being equivalent to that of malaria and TB (Hotez et al. 2006a).

Encouragingly, the NTDs are now receiving some of the attention their enormous burden warrants. As a result of impressive levels of advocacy (Hotez et al 2009b; Hotez et al. 2007; Molyneux et al. 2009; Molyneux et al. 2005; Engels & Savioli 2006), and the establishment of a range of public-private partnerships (Liese & Schubert 2009; Peters & Phillips 2004), co-ordinated international responses have been initiated towards at least 13 of the major neglected diseases (Brady et al. 2006; Hotez 2011; Hotez et al. 2007). The list of NTDs has since been extended to 17, and targets for their control and elimination by 2020 were recently set by the WHO (WHO 2012b), as well as by a number of international organisations from the public and private sectors and governments of the UK and USA (WHO 2012c). These and previous initiatives have made substantial contributions to reducing morbidity and mortality due to infectious disease in the developing world (UtCNTD 2014). Despite this, and the establishment of two journals dedicated to NTD research, published work on these diseases still lags well behind that of conditions with equivalent burden (Vanderelst & Speybroeck 2010).

Much more work is needed. To date, and with notable exceptions (Barry 2007), NTD control has focussed primarily on mass drug administration (Hotez 2009). Fears about the emergence of resistance, as well as the impact of vertical (disease-specific) treatment programmes on general health service provision (Marchal et al. 2009) means new treatments and vaccines are required (Hotez & Pecoul 2010), as well as research on effective ways to integrate them with other control measures (Parker & Allen 2011; Allotey et al. 2008). Observational studies also still have an important role to play, not least in describing the magnitude of the problem (Aagaard-Hansen & Chaignat 2010). Of growing interest, and a major focus of this thesis, is identifying the factors that influence co-occurrence of multiple NTDs in the same populations, as well as their relationship with the 'Big 3' infectious diseases (Fincham et al. 2003; Hotez et al. 2006b; Noblick et al. 2011). Basic research is needed on the biological interactions between these pathogens, but field-based studies can contribute to our understanding of the shared effects of social, structural and environmental forces that may act to shape co-endemicity (Gazzinelli et al. 2012). This information can contribute to the cost-effective targeting of control and monitoring activities that are integrated for multiple diseases (Hotez et al. 2006a; Molyneux et al. 2005; Clements et al. 2010a).

1.2. Study aims

In this thesis, I explore the factors that influence risk of infectious disease for people in a predominantly poor farming community in western Kenya. Using data from a community-based cross-sectional survey, I describe individual and household level prevalence and risk factors for infection with a wide range of endemic pathogens, including a number of NTDS, as well as HIV, TB and malaria. With particular attention to exploring social determinants, I attempt to show how shared risk factors for multiple pathogens might influence co-occurrence of infectious diseases within single households and across whole landscapes..

These general aims will be met through the following objectives:

1. Describe spatial heterogeneity in household risk for a range of infectious agents within a single area of western Kenya, with a focus on identifying spatial congruence that might indicate the existence of shared contextual or compositional effects;

2. Derive a locally appropriate index of household socioeconomic position and examine its effect on individual risk of infection for a range of locally prevalent pathogens with diverse transmission routes;

3. Show the utility of defining infection as a household-level process, particularly for neglected helminth infections, whilst highlighting important biases;

4. Explore shared risk factors that influence multiple infection with NTDs at the household-level.

1.3. Structure of the thesis

This thesis is structured as eight chapters that seek to contribute to each of these objectives. In the following chapter (**Chapter 2**), I extend and provide further evidence for the themes already introduced, with a focus on exploring vulnerability to NTD risk in sub-Saharan Africa. In addition, I describe the utility of considering infection at multiple scales, including the individual and household level. **Chapter 3** provides background on the methodology used for the collection and processing of data that forms the basis for the descriptive, exploratory and analytical work in this thesis. In **Chapter 4,** I describe the prevalence of both individual and household infection for a range of communicable diseases of people, and examine the extent to which risk of infection depends on residential location. Exploratory spatial approaches allow me to examine spatial overlap in infectious disease risk. In **Chapter 5,** multivariate statistics are used to derive a robust indicator of household socioeconomic

position (SEP), which is used to predict individual risk of infection for a range of prevalent infectious outcomes, as well as individual access to medical services. The use of zero-inflated binomial models as a means to derive household infection status on the basis of samples collected from individuals is introduced in **Chapter 6**, using the example of the soil-transmitted helminths (STH). These models are extended in **Chapter 7**, where I explore shared risk factors influencing household-level infection for multiple neglected helminth species. Finally, in **Chapter 8,** I summarise the findings in this thesis and discuss their wider applicability, particularly in light of the increasing need for improved methods for monitoring and surveillance and targeted disease control of the NTDs (as well as other important infectious diseases).

The thesis is written as a continuous document, with each chapter informing the work in the next, and I have sought to restrict repetition. Where appropriate, some additional background, more in-depth analysis and elaboration of results are provided in the Appendix. Since the "People, Animals and their Zoonoses" (PAZ) research project on which this thesis is based was very much a collaborative effort, I use "we" throughout, but all work in this thesis is my own.

Chapter 2

Background

2.1. The neglected tropical diseases

Neglected tropical diseases (NTDs) are the most common infections of the world's poorest people (Hotez et al. 2007). The conditions that the WHO (2010) lists for the NTDs are: 1) a proxy for poverty and disadvantage; 2) affect populations with little visibility and little political voice; 3) do not travel widely; 4) cause stigma and discrimination, particularly of girls and women and have an important impact on morbidity and mortality; 5) relatively neglected by research; 6) can be controlled, prevented and possibly eliminated using feasible solutions. Additional conditions could also include their ability to promote poverty through effects on child development, educational attainment, and adult productivity (Hotez et al. 2009b). The 17 'official' NTDs are shown in Table 2.1, together with several infectious diseases which meet the criteria listed above and for which advocacy is therefore still needed (Mableson et al. 2014).

The current status and what is known of the epidemiology of each of the 17 NTDs has been extensively reviewed (WHO 2012a; WHO 2012b). In the following sections I describe some of the factors that are pertinent to the group as a whole, and considerations related to the "infectious diseases of poverty" more generally.

Table 2.1. The neglected tropical diseases (modified from (Hotez & Pecoul 2010)).

Helminth diseases	**Bacterial diseases**	**Protozoan diseases**
Ascariasis*	Anthrax	Leishmaniasis*
Trichuriasis*	Bartonellosis	Chagas disease*
Hookworm*	Bovine tuberculosis	Human African trypanosomiasis*
Strongyloidiasis	Buruli ulcer*	Ameobiasis
Toxocariasis	Cholera	Giardiasis
Lymphatic filariasis*	Leprosy*	Balantidiasis
Onchocerciasis (river blindness)*	Leptospirosis	Toxoplasmosis
Loiasis	Relapsing fever	Trichomoniasis
Dracunculiasis (Guinea worm)*	Trachoma*	
Schistosomiasis*	Treponematoses (Yaws)*	**Ectoparasitic diseases**
Food-borne trematodiasis*		Scabies
Cysticercosis/taeniasis*	**Viral diseases**	Myiasis
Echinococcosis*	Dengue fever*	Tungiasis
	Japanese encephalitis	
Fungal diseases	Jungle yellow fever	
Paracoccidiomycosis	Rabies*	
Mycetoma	Rift Valley fever	
	Viral haemorrhagic fevers	

Conditions with * are specifically targeted by the WHO's roadmap for NTD control (WHO 2012a) (Note, ascariasis, trichuriasis and hookworm are classified by the WHO as a single condition; soil-transmitted helminthiasis)

2.2. Polyparasitism and co-endemicity

The NTDs are said to overlap and to cluster within poor communities (Aagaard-Hansen & Chaignat 2010), where they often co-occur with HIV, TB and malaria (Hotez, et al. 2006a). People living within these communities may therefore be at high risk for a very wide range of infectious diseases.

Polyparasitism, or co-infection with more than one infectious agent, is described as "the result of commonalities in ecological and environmental requirements, infection routes, host exposures, and susceptibility, as well as behavioural, sociological, and economic factors that enable co-occurrence of a multiplicity of parasite-host systems in time and space" (Gazzinelli et al. 2012). Identifying and addressing these 'commonalities' provides a potentially powerful means for the combined control of multiple pathogens (Ziegelbauer et al. 2012; Kolaczinski et al. 2007; Utzinger et al. 2003). Greater access to latrines, for example, would be expected to reduce the within-community prevalence of a number of NTDs, including soil-transmitted helminths (STHs) and gastrointestinal protozoa (Ziegelbauer et al. 2012; Schmidlin et al. 2013), *Schistosoma* spp. (Esrey et al. 1991), *Taenia* spp. (Jayashi et al. 2012; Mwape et al. 2012; Ekpo et al. 2008), *Chlamydia trachomatis* (the

cause of trachoma) (Montgomery et al. 2010; Emerson et al. 2004) as well as a range of diarrhoea-causing bacteria and viral infections (Genser et al. 2008). Similarly, malnutrition is a shared risk factor for a numerous infectious diseases in children (Rice et al. 2000; Black et al. 2010). Being underweight in childhood was estimated to account for 46% of the population attributable fraction (PAF) for lower respiratory tract infections in countries with high mortality rates, 49% for diarrhoeal disease, 45% for malaria and 34% for measles (Ezzati et al. 2003). In communities in which children are frequently underweight, or in which sanitation facilities are basic, and such communities tend to be one in the same (Dangour et al. 2013), it would be reasonable to expect co-infection with multiple pathogenic organisms to be extremely common. Indeed, polyparasitism is the prevailing state in people throughout the developing world (Petney & Andrews 1998; Raso et al. 2004; Utzinger et al. 1999; Guignard et al. 2000; Waikagul et al. 2002; Tchuem Tchuenté et al. 2003; Lwambo et al. 1999; van den Bogaart et al. 2012; Adegnika et al. 2010; Buck et al. 1978; Keiser et al. 2002; Yapi et al. 2014), as was probably the case in much of the 'developed' world before industrialisation (Gore & Custovic 2004; Le Bailly et al. 2012).

Whilst such commonalities in exposure in co-endemic areas are a necessary condition for co-infection, the establishment, replication, and persistence of each pathogen within the human host may also be a consequence of direct and indirect biological interactions between them (Cox 2001; Pedersen & Fenton 2007; Griffiths et al. 2011). Pathogen interactions within the individual can have important effects on transmission dynamics at a population-level (Gibson et al. 2010). Moreover, co-infected individuals are often at elevated risk of morbidity (Raso et al. 2004; Hürlimann et al. 2014). Individuals harbouring multiple soil-transmitted helminth (STH) infections (*Ascaris lumbricoides, Trichuris trichiura* and hookworm), for example, often have the most intense infections and are at greatest risk of anaemia and nutritional deficiencies (reviewed in (Pullan & Brooker 2008)).

Despite this, and despite the likely frequency of co-infection in developing countries, field-based studies of human polyparasitism are still relatively rare (Steinmann et al. 2010; Hürlimann et al. 2014), a situation that is likely to have consequences for correctly estimating the morbidity associated with single conditions (Pullan & Brooker 2008).

2.2.1. Relationship with the 'Big 3'

There is a growing evidence base that suggests important interactions between NTDs, particularly the helminthic infections, and HIV (extensively reviewed by Brown et al. (2006); Harms & Feldmeier (2002); Webb et al. (2012); and Karp & Auwaerter (2007b)).

Some of these effects are species-specific, and relate to the mechanical effects of infection on HIV susceptibility. *Schistosoma haematobium*, for example, which occurs mainly in sub-Saharan Africa, causes friable, genito-urinary lesions which may increase the rate of sexual transmission of HIV (particularly to females), and is likely to be a major co-factor for transmission of the virus in co-endemic areas (Kjetland et al. 2006; Mbabazi et al. 2011; Ndeffo Mbah et al. 2013).

It is also well known that helminth infection can lead to polarisation of the immune response to one that is predominantly Th2, rather than Th1, mediated (Fincham et al. 2003; Else et al. 1994; Gaze et al. 2012). Broadly, Th1 cells (producing cytokines such as IL-2 and IFN-gamma) promote cell mediated immune responses, while Th2 cells produce cytokines that mediate the humoral (antibody) response (Oosterhout & Motta, 2005). It has been suggested that the Th2 'priming' associated with helminth infection may increase the risk of HIV transmission: a Th2 dominated immune profile can result in dampening of the Th1 response, which is necessary for the effective control of intracellular infections, and particularly the suppression of CD8+ mediated viral clearance (Kullberg et al. 1992; Actor et al. 1993). Moreover, the successful establishment of HIV once it has entered the systemic circulation relies on the presence of activated immune cells: resting T cells and undifferentiated monocytes typically have low susceptibility to HIV infection (Harms & Feldmeier 2002; Card et al. 2009; Koning et al. 2005; Bégaud et al. 2006). Th2 activation of both T cells and peripheral mononuclear cells as a result of helminth (and other) infection may result in the up-regulation of co-receptors on cell surfaces that can make them highly susceptible to HIV (Maggi et al. 1994; Bentwich et al. 1998; Secor et al. 2003). Parasitism by helminths is known to result in T cell activation (Kalinkovich et al. 1998; Chachage et al. 2014) and peripheral mononuclear cells from individuals parasitised by helminths have been shown to be more susceptible to HIV infection than those from individuals without (Shapira-Nahor et al. 1998; Gopinath et al. 2000). Moreover, the presence of gastrointestinal helminths, and subsequent mucosal damage, may allow increased levels of bacterial translocation (George et al. 2012) and subsequent immune stimulation, with a resultant increase in the risk of HIV establishment following exposure (Douek 2007; Pantaleo et al. 1993).

In addition to these immunological effects on susceptibility, there is increasing evidence that HIV-helminth co-infected individuals may be more infectious. Mothers with HIV helminth co-infections, for example, were found to be more likely to transmit HIV to their offspring than HIV infected mothers without helminths (Gallagher et al. 2005), although deworming during pregnancy was not found to affect the risk of transmission in a randomised controlled

trial (Webb et al. 2011). Several studies have shown a trend for higher HIV viral titres in helminth co-infected individuals (Mulu et al. 2013; Wolday et al. 2002; Bentwich et al. 2000) which tend to decline following anthelmintic treatment (Wolday et al. 2002; Walson & Stewart 2007; Modjarrad & Vermund 2010; Sangaré et al. 2011; Mulu et al. 2013; Bentwich et al. 2000), although this effect is not consistent across all studies (Modjarrad et al. 2005; Nielsen et al. 2007; Assefa et al. 2009; Elliott et al. 2003; Hosseinipour et al. 2007). Experimental studies involving SIV infection in macaques reported higher viral loads in *Schistosoma mansoni* co-infected animals (Chenine et al. 2005), which were also more susceptible to infection with the virus than non-parasitised controls (Chenine et al. 2008).

There has been surprisingly little work to explore the consequences of helminth co-infection on HIV transmission dynamics at the population level (Gibson et al. 2010; Mushayabasa & Bhunu 2011). Given the geographical overlap of HIV and several helminth NTDs, and the ubiquity of the latter in the areas that are currently the focus of the HIV pandemic (Hotez et al. 2006a), such work is urgently needed. Indeed, some authors have suggested the effects of parasite co-infection on both HIV infectiousness and susceptibility may be a major factor driving the force of on-going HIV epidemics in sub-Saharan Africa (Sawers & Stillwaggon 2010; Cuadros et al. 2011).

In addition to consequences for HIV transmission dynamics, it has been suggested that the immuno-modulatory effects of helminth infection described above may also impair the response to *Mycobacterium tuberculosis* (an intracellular infection causing tuberculosis which also relies on Th1 mediated clearance) and therefore impact on infection, progression to active disease, and resulting pathology (Elias et al. 2007; Resende Co et al. 2007; Borkow et al. 2001), as well as on the efficacy of BCG vaccines (Elias et al. 2001). However field-based evidence remains limited and contradictory (Elias et al. 2006; Elias et al. 2001; Chatterjee et al. 2014; Tristão-Sá et al. 2002), and recent experimental studies were unable to demonstrate an impact of helminth infection on *M. tuberculosis* progression in cotton rats (Hübner et al. 2012), a finding repeated in other rodent models (Erb et al. 2002; Frantz et al. 2007). In wildlife populations, however, nematode infections were shown (on the basis of a dynamic model using empirical data) to have important impacts on *Mycobacterium bovis* transmission, to the extent that the authors concluded that *M. bovis* would be unlikely to have invaded a population of African buffalo in the absence of helminth co-infection (Ezenwa et al. 2010).

There is substantial overlap of *Plasmodium falciparum* malaria and helminth infections in sub-Saharan Africa, and co-infection is ubiquitous in many areas (Brooker et al. 2006c).

Whilst infection with both *Plasmodium* spp. and helminths may cause anaemia (although this varies between helminth species (Smith & Brooker 2010)), a recent meta-analysis was unable to show an additive effect of malaria-helminth co-infection (Naing et al 2013). The immunological effects, and epidemiological consequences, of co-infection are also uncertain (Mwangi et al. 2006). A range of observational studies have shown positive statistical associations between *P. falciparum* and helminth infection (reviewed in Kobylinski et al. (2014)), but relationships with clinical malaria tend to be quite variable. Brutus et al. (2007 & 2006) report that malaria parasite densities were reduced in individuals treated for *Ascaris*, and people infected with this parasite have been found to have reduced risk of cerebral malaria (Nacher et al. 2000). By contrast, Le Hesran et al. (2004) reported that the prevalence of *Ascaris* infection was higher in those with severe malaria than those without (although these findings have been questioned (Nacher 2005)), and several other authors suggest increased risk of clinical malaria in helminth co-infected individuals (Spiegel et al. 2003; Sokhna et al. 2004). In a meta-analysis of helminth-*Plasmodium* interactions in mice, Knowles (2011) also reported substantial heterogeneity in effects across experimental studies.

In the next section, I describe some of the available control strategies for the major NTDs, but to date the primary approach has been the use of one or all elements of a "rapid impact" package of preventive chemotherapy (Hotez et al. 2007). These treatments can be expected to reduce morbidity associated with the NTDs they target (Bundy et al. 2013), but the evidence described above suggests the pharmacologic removal of the these pathogens may also have important impacts on co-infecting agents (Knowles et al. 2013). Clearly these effects need to be taken into account in cost-effectiveness and impact assessments, and much more work is needed to describe and explain them (Fenton 2013; Wiria et al. 2013; Webb et al. 2012).

2.3. Control of NTDs

The very low cost of highly efficacious pharmaceutical agents (which are also very often donated by manufacturers (WHO 2012c)), as well as their very high safety profile, means preventive chemotherapy, or mass drug administration (MDA), has become the cornerstone of NTD control (WHO 2006b; WHO 2012a). Between 2006 and 2012, 589 million NTD treatments for soil-transmitted helminths (STH), schistosomiasis, lymphatic filiariasis, onchocerciasis and trachoma were provided (Wouters et al. 2014). Targets were set for the treatment of 75% of all children in endemic areas with albendazole or mebendazole (for the

STH) by 2010 (WHO 2012a), and country-wide school-based deworming programmes for the STH and schistosomiasis are underway in Kenya (Mwandawiro et al. 2013). Worldwide, these interventions have undoubtedly had a major impact on reducing morbidity (Guyatt et al. 2001; Kabatereine et al. 2007; Bundy et al. 2013; Ramzy et al. 2006), and provide a rapid, cost effective and relatively simple means for control without the requirement for major infrastructural changes, which may be difficult to achieve in some settings (Brady et al. 2006; Linehan et al. 2011).

However, the focus on chemotherapy is not without criticism (Humphries et al. 2011; Utzinger et al. 2011; Campbell et al. 2014), and for several of the NTDs could be considered to represent a 'fire fighting' approach, rather than one that tackles root causes. For the environmentally transmitted helminthic NTDs (such as STH and schistosomiasis), for example, chemotherapy targets only the population of parasites that exist in infected individuals. Following treatment, people living in areas with poor sanitation and inadequate waste management are likely to become rapidly re-infected by the large population that may remain in the environment (Yap et al. 2013; Quinnell et al. 1993; Jia et al. 2012). This 'bounce-back' suggests anthelmintic coverage would need to be extremely high, and be administered regularly, to achieve the elimination of helminth transmission (Anderson et al. 2014). A move from targeted drug administration to school age children, who are (generally) at greatest risk of morbidity, and therefore benefit most from treatment, to mass drug administration applied to whole communities is therefore increasingly proposed (Mwinzi et al. 2012). This approach has been successful for the control for lymphatic filariasis (Wamae et al. 2006; Sodahlon et al. 2013) and onchocerciasis (Traore et al. 2012). However, mass drug administration programmes to control gastrointestinal nematodes in livestock have seen the emergence of profound anthelmintic resistance (Wolstenholme et al. 2004). Whether similar levels of anthelminthic resistance will emerge among the important human gastrointestinal nematodes (the STH and *Strongyloides stercoralis*), which probably infect more than 1 billion people worldwide (Pullan et al. 2014b), remains uncertain (Vercruysse et al. 2011; Vercruysse et al. 2012). However, the number of licensed treatments for the helminthic NTDs are currently very limited (Keiser & Utzinger 2008), and therefore the prudent use of currently available drugs is an important consideration (Lustigman et al. 2012).

Structural interventions that seek to reduce people's exposure to NTDs would be one such approach to minimise reliance on pharmaceutical interventions. For example, two recent meta-analyses showed that access to latrines substantially reduced an individual's risk of

infection with the STH (Strunz et al. 2014; Ziegelbauer et al. 2012), and that the use of treated water sources was also protective (Strunz et al. 2014). A recently published longitudinal study from China examining the integration of biomedical with development-based interventions, such as improved sanitation, piped water provision, mechanisation of agriculture (to reduce zoonotic exposure from draft animals) and environmental modification resulted in long term reductions in the prevalence of schistosomiasis (Chen et al. 2014). Indeed, it is likely that economic development and associated advancements in water and sanitation are responsible for many of the declines in STH and schistosomiasis prevalence observed in some endemic countries (Knopp et al. 2013; Pullan, Gething, et al. 2011a; de Silva et al. 2003), as was the case in the developed world (Spiegel et al. 2010).

Development-based control is not included on the WHO roadmap for NTD elimination (WHO 2012a), but, in the interests of long term sustainability, and particularly for the neglected helminth infections, much more work is needed to examine the effectiveness of approaches that integrate (and potentially replace) vertical, biomedical NTD control with structural interventions such as the strengthening of primary health care provision (Mecaskey et al. 2003; Glenngård & Maina 2007) as well as sanitation and water projects (Singer & de Castro 2007; Utzinger et al. 2011). Research on the effectiveness of these interventions for NTD control has lagged well behind that in the biomedical realm (Allotey et al. 2008; Reidpath et al. 2011; Manderson et al. 2009), to the extent that some authors have suggested a 15% "social offset" for any research funding towards NTD drug or vaccine development (Spiegel et al. 2010). Greater information on the costs and benefits for NTD control of interventions that seek to address the impact of social inequalities (or the 'social determinants of health') is required for informed policy making and resource allocation.

2.4 The social determinants of the NTDs

At around the same time that the neglected tropical diseases were coming to the general attention of a wider scientific and policy making community, another 'neglected' epidemiological focus was also emerging (or re-emerging). In 2008, the WHO reiterated the findings of Sir Michael Marmot and the Commission on Social Determinants of Health that, in order to improve global health, action was needed to reduce social and economic inequalities across a broad range of policy areas (CSDH 2008; Bell et al. 2010). Consideration of 'contextual' effects, including poverty, government policy, and the practices of the private sector, and their interaction with individual level factors in disease causation, were, and continue to be, in the ascendance. This increasing focus on the distal

determinants of health followed (temporally, at least) much soul searching by many in the epidemiological community about the increasingly individualistic (or proximal) focus of the discipline (Krieger 1994; Pearce 1996; Rothman et al. 1998; Shy 1997; Susser & Susser 1996; Diez-Roux 1998).

By their very definition, the NTDs are inherently linked to social determinants of health (SDH), particularly household-level poverty (Conteh et al. 2010), but also poor infrastructure, the effects of migration, conflict, and gender inequality (Aagaard-Hansen & Chaignat 2010; Alsan et al. 2011). And yet, the NTD and SDH (or social epidemiological), communities were said to have "passed in the night" (Spiegel 2011), with minimal interaction between the two. Indeed, Cohen et al. (2007) report that of the 279 publications that used the term "social epidemiology" between 1966-2005, only 15 (5.4%) investigated infectious outcomes, and these were primarily sexually transmitted infections (STDs) in high income countries.

The apparent disconnect between social epidemiology (which has emerged as a discipline to study the socio-environmental determinants of health (McLaren & Hawe 2005)) and NTD research may have more to do with a lack of a shared language than a lack of 'social' research on neglected diseases (Spiegel 2011). Indeed, risk factor studies that explore NTD infection at the individual level are now very commonly multilevel, including a range of effects, such as socioeconomic status, income, and maternal education, operating at a distal, household-level, in addition to individual factors such as age, sex, behaviour and nutrition that act at the individual (proximal) level (Schüle et al. 2014; Riess et al. 2013; Raso, Vounatsou, Gosoniu, et al. 2006a; Halpenny et al. 2013; Schmidlin et al. 2013; Nundy et al. 2011; Pullan et al. 2008). Yet few authors describe these in 'contextual' terms, or as distal and proximal determinants. Similarly, the two recent meta-analyses already described and reporting on WASH, the combined effects of water, sanitation and hygiene on soil-transmitted helminth risk, clearly demonstrated the benefits of these contextual interventions on the health of individuals in multiple populations (Ziegelbauer et al. 2012; Strunz et al. 2014). But neither mention "contextual effects" or "social determinants".

Moreover, the impacts of NTDs, particularly the soil-transmitted helminths, on early childhood development, one of the major social determinants of health (Davey-Smith 2003), have been reasonably well studied and were subject to a recent Cochrane review (Taylor-Robinson et al. 2012). Whilst the outcomes were equivocal (and controversial (Bundy et al. 2013)), they frame the debate for school-based deworming as an intervention that is explicitly targeted towards the determinants of inequality (Bundy et al. 2013; Hotez et al.

2009b). Work is also underway to explore behavioural factors that influence uptake of control strategies, particularly MDA (Phongluxa et al. 2014), as well as the impacts of education on hygiene behaviours and STH risk (Gyorkos et al. 2013; McManus et al. 2014). Finally, spatial analysis and disease mapping is increasingly being used to examine how broad-scale social inequalities, such as poverty, access to sanitation, and water supply, may shape observed geographic heterogeneities in NTD risk (Pullan et al. 2014a; Soares Magalhães et al. 2011a).

It would therefore be difficult to argue that the social determinants of health are not being considered by those working on NTDs, even if not always explicitly so. But a much greater integration of the two communities is likely to contribute to improved understanding of, and ability to control, the infectious diseases of poverty (Aagaard-Hansen & Chaignat 2010).

2.5. Household vulnerability to infectious disease

The household is at the centre of the transmission process for many infectious diseases, and particularly those that are the result of poor sanitation, contaminated water supply and unhygienic practices (Cairncross et al. 1996; Bethony et al. 2001; Walker et al. 2011; Lescano et al. 2009; Khan 1982; Simpson 1952), as well as for some vector-borne diseases (Wahyuni et al. 2004; Coreil et al. 2000; Gürtler et al. 1998; Levy et al. 2007). In contextual terms, the household is also the first level that links individuals to higher-order social effects at the community, regional or national level (Dewalt & Pelto 1985) and as such there has been a long tradition of considering health as a household-level process in the social sciences (Popkin 1982; Coreil et al. 2000). The “household production of health”, for example, describes the "dynamic behavioural process through which households combine their (internal) knowledge, resources, and behavioural norms and patterns with available (external) technologies, services, information, and skills to restore, maintain and promote the health of their members" (Berman et al. 1994). Processes that influence the allocation of resources to health, and therefore to a household’s vulnerability and resilience to disease, could be expected to contribute to variation in disease burden within a single community. Bates et al. (2004b) explored some of these processes in relation to HIV, malaria and TB, and considered effects such as poverty, nutrition, livelihoods, gender, education, and religion as being particularly important in explaining household vulnerability.

Infectious disease could also be considered as a household-level outcome within an explicit risk-based framework (Coreil et al. 2000). In risk analysis, ‘risk’ combines the probability of an adverse event occurring, with the probabilities of a set of consequences if it does (SRA

2014). Hence, we could consider the likelihood that an infectious disease could be introduced to a particular household, and the consequences of that introduction in terms of onward spread and persistence. A similar, although perhaps less applied, definition could also be given to vulnerability, as the "exposure to contingencies and stress and difficulty in coping with them" (Chambers 1998). The NTDs, and particularly the neglected helminths, are often highly ubiquitous in endemic communities (Aagaard-Hansen & Chaignat 2010). However, through their effects on reduced productivity, stigma or health expenditure, infections may have the greatest impact in the poorest households (Xu et al. 2003; Whitehead et al. 2001; Boelaert et al. 2010; Conteh et al. 2010). These families may put the lowest priority on health expenditures, or have least access to health care, and therefore infection with NTDs, which tend to cause chronic conditions that often become progressively worse without treatment, may persist for longest in individuals living in the poorest households (Singer & Ryff 2007). Clearly, these diseases also do not occur in isolation and the economic consequences are likely to be felt by all members of a single family (Perera et al. 2007; Wyss et al. 2004; Chuma et al. 2006). Moreover, and given that many of these infectious agents, and particularly the environmentally transmitted helminths, tend to cluster within households (Walker et al. 2011; Cairncross et al. 1996; Brooker et al. 2006a; Forrester et al. 1990), all individuals within an 'infected' household may be at high risk of infection once the agent is introduced to the domestic environment. The quantitative study of factors that influence both the likelihood of household-level infection *and* the consequences of infection (which might be related to within-household transmission, duration of infection or to the economic consequences) could therefore be usefully integrated into a risk-based surveillance or control programme that seeks to target these high risk households within endemic communities (Stärk et al. 2006).

2.6. Household infection as a substantive outcome

Considering household-level effects on individual-level outcomes is not new, and multilevel models are now widely used to explore conditions that operate at multiple scales (including the household) in infectious disease research (Diez-Roux 2000; Diez-Roux & Aiello 2005; Rothman & Greenland 1998). However, current applications of multilevel models tend to report the effects of predictors (at multiple levels) operating on infectious outcomes at the individual-level only. Whilst covariate-outcome relationships are used (very valuably) to explain some of the reasons why infectious agents may cluster within particular households, few authors have quantitatively explored the factors that influence the dichotomy of the presence or absence of infection within the household itself (but see Olsen et al. 2001; Root

et al. 2013; Marcia Caldas de Castro et al. 2007; Redlinger et al. 2002; Trönnberg et al. 2010), Quantitative studies that use household infection as a substantive outcome could potentially reveal patterns regarding the dynamics of infection within an endemic landscape. For example, it is a common observation that individuals living in the poorest households are at greatest risk of STH infection. This does not necessarily mean the poorest households are more likely to have *any* infection: a positive relationship with poverty could be observed if all households, rich and poor, are at equal risk of infection, but if *more* people are infected in poorer households. Hence, the reverse of the ecologic fallacy, or the bias that sometimes arises when inference on an outcome measured at the individual level is made based on aggregate (ecological) data (McLaren & Hawe 2005; Morgenstern 1995), is also true.

Species distribution (or ecological niche) models are increasingly being used to explore the impacts of landscape heterogeneity on the geographic distribution of pathogenic agents (Peterson 2006). The presence or absence of a pathogen at the household-level may be a particularly useful outcome to explore social and environmental landscape effects that influence the spatial distribution of disease and its limits. To that end, household level infection has previously been used as an outcome to explore small-scale spatial heterogeneities in risk for a range of pathogens (Zhou et al. 2013; Halpenny et al. 2013; Swaminathan et al. 2012; Handzel et al. 2003; Booth et al. 2004; Saathoff et al. 2005).

Considering infection status at different levels is likely to be particularly useful when interventions are to be targeted at those levels (Susser 1994). Focussing disease screening on those households that are most likely to be infected might have particular utility for active surveillance (Stärk et al. 2006), and towards the end stages of disease elimination programmes (WHO 2002). Targeting the highest risk households (in terms of probability of infection or the consequences of infection) could potentially also have a role in community based mass drug administration programmes for prevalent helminth species (such as the STH). In particular, providing treatment only to those households that would benefit most from it could be an effective method to reduce selection pressure for anthelmintic resistance by promoting the concept of refugia (van Wyk 2001; Kenyon et al. 2009).

An obvious question then arises; how does one define household ‘infection’? A straightforward definition would be the presence of at least one infected individual, and this definition has previously been used in infectious disease studies to explore spatial heterogeneity in household infection risk. Such an approach is also very common in veterinary medicine, where ‘herd prevalence’, or the probability that a randomly selected herd contains at least one infected animal, is commonly used for disease surveillance and

control activities (Christensen & Gardner 2000; Jordan & McEwen 1998). In that discipline, herd infection is also often used as an outcome in logistic regression models (for example Bae et al. (2013); Bronsvoort et al. (2001). Here, the group represents the sampling unit, the unit of interest, and the unit of analysis (McDermott & Schukken 1994).

Of course, an, important consideration in using an ecologic (aggregated) outcome, and perhaps part of the reason why household level infection is used relatively rarely in human epidemiology, is that is an inherently reduced information state (Morgenstern 1995; Greenland 2001). This is likely to have particular consequences for inference when the presence or absence of infection in the group is defined on the basis of imperfect diagnostic tests applied to individuals. In such cases, the probability of correctly identifying the group as infected (or uninfected) will depend on the within-group prevalence (Christensen & Gardner 2000). When conditions impact upon both within- and between-group prevalence (i.e. the probability of individual infection and the probability of household infection), there can often be substantial bias in the effects of these conditions on group level infection if the sensitivity of the sampling process used to define group status is not adequately accounted for (Kery & Schaub 2012; Tyre et al. 2003; MacKenzie et al. 2002). This fact tends to be ignored when modelling herd level outcomes in the veterinary literature (de Glanville et al *in prep*.), but means that the outcome being modelled is at best a measure of group-level detection rather than infection.

These issues could make disentangling the effects that influence household-level infection from those that influence individual-level infection a challenging venture. However, the relatively recent emergence of zero-inflated, or two-part models that allow the partitioning and joint modelling of effects at different levels (Wenger & Freeman 2008; Martin et al. 2005; Stryhn & Christensen 2014), provide a possible solution.

2.7. Zero inflated models

Zero inflated models consider data as consisting of two components: a zero component and a non-zero component, which are modelled separately (Stryhn & Christensen 2014). The zero component consists of systematic zeroes, whilst the non-zero component includes observations that could (but do not necessarily) provide a positive count. In an infectious disease application, the first of the two components might consider the probability that the abundance of infection is greater than zero (i.e. model a prevalence outcome), while the second component considers the mean abundance given that it is greater than zero. The first component could be modelled with a logistic regression, and the second using linear or non-

linear regression (Stryhn & Christensen 2014). As such, zero inflated models can be considered to be a two part mixture distribution, with the output from the first part "inflating" the output from the second via a Bernoulli distribution (of 0 or 1's) (Zuur et al. 2012).

Importantly, the separate models can incorporate both random and fixed effects at the level of each component, and can therefore be used to test competing hypotheses in the same system. Kristoffersen et al. (2013), for example, used a zero-inflated approach to model factors that influence introduction of sea lice into Chilean salmon farms *and* the factors that influence onward spread and persistence on infected farms. The first outcome was presence or absence of sea lice (modelled with a logistic regression), the second was the abundance of sea lice (modelled using gamma regression). A similar application of a zero inflated model was used to model individual infection with *Schistosoma haematobium* in school children in Malawi (Chipeta et al. 2013). Here, a logistic regression model considered the probability of infection, whilst a negative binomial regression was used to describe infection intensity. These authors used these two part models to describe a range of factors that influence infection and intensity of infection, and were able to show that whilst males were less likely than females to be infected by *S. haematobium*, they tended to have more intense infections. Given that the probability of identifying *Schistosoma* spp. in a faecal or urine sample is related to the number of worms present in a parasitised individual (Krauth et al. 2012), the confounding effect of infection intensity on detection may have made such an effect difficult to observe without zero-inflation.

These models have also been used to explore infection at the level of the household. Root et al. (2013) used a zero inflated longitudinal Poisson regression to examine factors that influence both the likelihood that cholera occurs in a household (or collections of patrilineal households called *baris*) as well as the number of cases that occur over time. They found that baris with lower socioeconomic status were both less likely to be infected by cholera, and also to have lower rates of infection.

Zero inflated models are also beginning to be used for disease surveillance and for risk mapping (Vergne et al. 2014; Benschop et al. 2010; Soares Magalhães, Biritwum, et al. 2011b; Vounatsou et al. 2009). Here they have particular utility in that they allow identification of presence or absence of infection, whilst controlling for the sensitivity of the underlying sampling process, or the probability of detection given presence (i.e. a two part model, one of which describes probability of infection, or that the abundance is greater than zero, and a second part that describes probability of detection given presence). It is therefore

quite striking that whilst there are currently only a handful of studies that have used zero inflated models for this purpose in epidemiology, they have become almost *de rigour* in ecology over the past decade to control for imperfect detection in species surveys (Guillera-Arroita et al. 2014; MacKenzie et al. 2006). Indeed, the first description of the application of these models for that purpose (MacKenzie et al. 2002) has now been cited nearly 1000 times.

2.8 Exploring vulnerability to infectious disease in rural western Kenya

In this thesis, I adopt a broad eco-epidemiological approach to understanding disease risk within a single community by exploring infectious outcomes and their determinants at a range of scales (Krieger 1994; Susser & Susser 1996). I focus on the contextual and compositional factors that may predict both individual- and household-level risk of infection, paying particular attention to those shared effects that influence risk for multiple pathogens. I therefore consider individual- and between-household prevalence to be distinct outcomes which, when considered in parallel, can potentially provide greater insight into the ecology and epidemiology of infectious disease within an endemic landscape. In particular, by examining both individual and household risk, we can begin to disentangle some of the factors that may influence the introduction of a pathogen into a household, and those that influence its spread or persistence following introduction. This approach recognises that individuals may be exposed to infectious agents both outside and within the domestic environment (Cairncross et al. 1996), that factors influencing these transmission events may have heterogeneous effects in either environment, and that transmission in each environment may have variable importance for different pathogens.

Whilst this thesis aims to examine risk of infection for a wide range of endemic infectious agents, I focus in particular on neglected helminths. These parasite species are typically highly prevalent within endemic communities (Pullan et al. 2014b), but, as environmentally transmitted agents, their distribution may be conditioned by a range of social and environmental factors (Brooker et al. 2006b; Bethony et al. 2006; Scott 2008). Moreover, transmission of intestinal helminths is well known to be centred at the household level (Otto 1931, Bundy & Medley 1992), where they tend to be highly aggregated (Woolhouse et al. 1997). Cairncross et al (1996), in introducing the paradigm of domestic and public 'domains' of disease transmission (i.e. within and outside the household), suggested measurement of household level clustering as a simple means with which to assess the importance of transmission within an infected household. The use of multilevel models to partition individual variation in risk of infection between households, and to explore some of the

household level factors that may explain the identified variation, will be a major theme of this thesis. The presence of spatial clustering of infection, particularly in household-level infection, may also suggest that between-household (public) transmission of infection is important in particular geographic areas, and suggests the existence of spatially heterogeneous contextual or compositional effects that influence rates of household exposure and/or susceptibility to infection. A second major focus of this thesis will therefore be the examination of clustering in household-level infection, and the factors that may explain it.

A representation of within and between-household clustering of infection within a single landscape is given in Figure 2.1. On the basis of this simple framework, we might expect there to be factors that operate on individuals living in particular households that contribute to higher than average rates of within-household (domestic) transmission once a pathogen has been introduced (resulting in the red cluster in Figure 2.1). Similarly, we could expect factors to influence the transmission of infection *between* households due to transmission in the public domain. Where these factors result in higher than average rates of between household transmission within particular areas, we could expect spatial clustering in household-level infection (resulting in the blue cluster in Figure 2.1). Considering helminth infection, Table 2.2 lists some of the factors that could be hypothesised to contribute to higher than average rates of within-household transmission, and those that may influence rates of between-household transmission. Clearly, many of the factors overlap, and a final major focus of this thesis will be the use of hierarchical, zero inflated regression approaches to try to disentangle some of these effects on within- and between-household transmission, with a focus on neglected helminth infections.

The sections below provide further detail on factors influencing transmission of neglected helminths, particularly the soil transmitted helminths (STH) (hookworm, *Ascaris lumbricoides* and *Trichuris trichiura*), in each environment. Additional detail is given in Chapter 6 and 7.

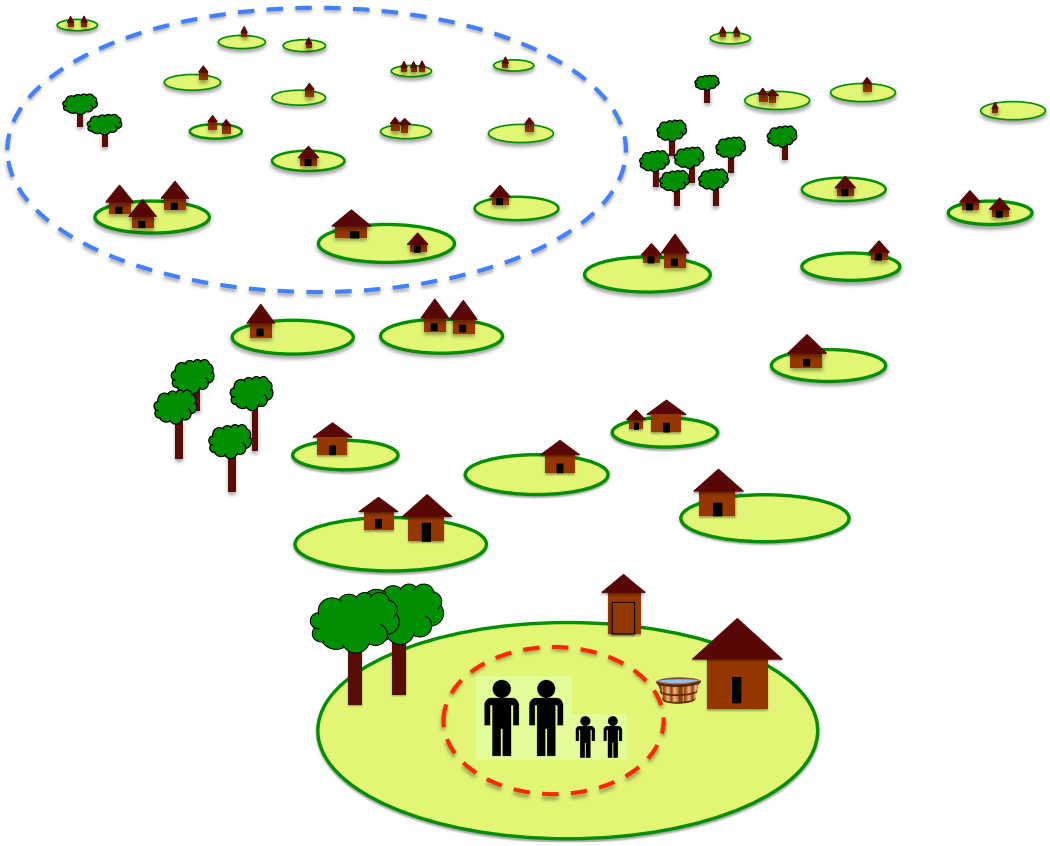


Figure 2.1. Clustering of infectious disease risk at two scales within an endemic landscape: within-households (red) and between-households (blue).

Table 2.2. Hypothesised conditions influencing within- and between-household transmission for neglected helminth species.

Within-household	Between-household
Environmental	***Community level*** *
Local vegetation cover	**Environmental**
Temperature	Vegetation cover
Precipitation	Temperature
Microclimate	Precipitation
Social/economic	Land use
Socioeconomic status	**Social/economic**
Sanitation	Socioeconomic status
Access to medical care	Level of sanitation
Education	Water supply
Behavioural practices e.g. hygiene	Public health interventions
Demographic	Population density
Age/sex composition	**Other**
Household crowding	Prevalence of co-infecting pathogens
Others	Nutritional status
Nutritional status of household members	***Household-level*** #
Immunological state	Age/sex composition
Presence/prevalence of co-infections	Livelihoods
Household hygiene e.g. presence of livestock	Number of people
	Socioeconomic status

* Factors operating at the community level that might be expected to influence between household prevalence of infection; # Factors operating on individual households that would be expected to influence the vulnerability of a particular household to disease introduction.

2.8.1 Within-household transmission

Once a pathogen has been introduced into a household, factors related to individual exposure and susceptibility to infection will influence the extent to which it may spread within the domestic environment. Infection with the soil transmitted helminths results from exposure to infectious larvae or eggs, and will therefore depend on the degree to which the environment may be contaminated, and the ability of the larvae and eggs to survive (Brooker et al. 2006b). Levels of environmental contamination could be expected to be influenced by the presence or absence of a latrine, which is likely to be determined by the socioeconomic status of the household (Jenkins & Curtis 2005). Socioeconomic effects, particularly the education of household members, are also likely to influence defaecation practices, as well as the extent to which potentially infected household members may be able to access medical care and anthelminthic treatments (more detail is given in Chapter 5). Demographic factors, such as household crowding, could also be expected to influence the density of infectious

agents in the household environment. Moreover, the age structure of the household may be important: young children tend to have the most intense infections with *A. lumbricoides* (Walker et al. 2011) and *T. trichiura* (Bundy & Cooper 1989), and may therefore be most important for seeding the household environment (Killewo et al. 1991), whilst adults may be more important for hookworm (Brooker et al. 2004a).

The extent to which larvae and eggs survive in the environment, and conditions required for maturation, will depend on the household microclimate (Brooker et al. 2006b). This will be influenced by macroclimatic conditions of the area in which the household is located, particularly temperature, humidity and precipitation. At the micro-level, it is likely to be influenced by the presence and type of vegetation and, potentially, the presence of livestock within the household environment which may impact upon soil conditions (Wang et al. 2012).

Factors influencing individual susceptibility to infection following exposure are likely to include an individual's immunological state, which may be heavily impacted by their nutritional status (Hughes et al. 2004). As already described (section 2.2), co-infecting agents can also result in immunomodulation that may impact individual susceptibility to infection.

2.8.2 Between-household transmission

A similar set of conditions as those that operate on individual exposure and susceptibility within infected households could also be expected to influence rates of between-household transmission. If public transmission can be assumed to occur when household members are exposed to infectious agents outside the home, or due to non-household members with patent infections entering a household, the probability of exposure will depend, in part, on the prevalence of a particular agent in the local community. Community level effects on the probability of exposure to the STH will therefore include environmental conditions, local vegetation cover and land use which, as already described, will influence the survival and development of eggs and larvae in the environment. The average socioeconomic status of the local community, impacting upon the proportion of households with adequate sanitation, frequency of open defaecation, and local water supply would also be expected to impact upon levels of contamination in public (i.e. non-household) areas of the community. Clearly, the extent to which public health interventions are in operation, and particularly mass drug administration for the STH, will also impact upon the community prevalence of infection.

For a specific household, the probability that an STH species is introduced will be influenced

by the extent to which household members may be exposed in public domains. Hence the presence of school attending children may increase risk of pathogen introduction. Moreover, particular livelihoods, such as agricultural work and repeated soil exposure, may also be an important risk factor for the soil transmitted helminths (Ensink et al. 2009). These factors will be related to the socioeconomic position of the household, which may also influence the degree to which household members are 'connected' to the local community in terms of the time they spend in public areas and the distances they travel.

Chapter 3

Study design

3.1. Introduction

All data described in this thesis were collected as part of the 'Epidemiology of zoonoses in livestock and their keepers' (PAZ) project, a multidisciplinary, 'one-health' study (Zinsstag et al. 2011) involving the University of Edinburgh (UoE), the International Livestock Research Institute (ILRI), and the Kenya Medical Research Institute (KEMRI), and funded by the Wellcome Trust, UK (Grant code: 085308). The main focus of the project was zoonotic disease in a sympatric population of animals and people in western Kenya. Substantial work to explore the distribution and determinants of the risk of zoonoses in these populations has already been undertaken (Thomas 2013; Cook 2014).

The work described in this thesis examines the epidemiology and ecology of a wide range of helminthic, protozoal, bacterial and viral infections of people, of both zoonotic and non-zoonotic origin. This short chapter provides the background on the collection, processing and testing of the biological material that forms the basis for all analysis in the following chapters. Given the focus on infectious diseases of humans, I only describe those parts of the PAZ study design that relate directly to people.

3.2. Study area

The study population was a mixed-farming community in western Kenya, in an area broadly representative of the wider Lake Victoria Crescent ecosystem which spans Kenya, Uganda and Tanzania. This region is characterised by rainfall and temperatures that are typically sufficient for two cropping seasons per year, and in which a range of subsistence and cash-crops are grown by the majority of rural households (Waithaka et al. 2002). Livestock, specifically local breeds of cattle, sheep and goats, and smaller numbers of pigs, are integrated with crop production in this mixed system through the use of manure as fertiliser, cattle as draft power, and crop surplus and residues as animal feed (Conelly & Chaiken 2000).

The study area was an approximately 3200 square kilometre zone defined by a semi-circle with a 45km radius emanating from the county town of Busia (in which the field station was situated) on the Kenyan border with Uganda (Figure 3.1). This area comprises a human population of 1.4 million people (Open Kenya 2014), and has among the highest population densities in Kenya.

It is known that the human population under study is endemic for soil-transmitted helminths and schistosomiasis (Brooker et al. 2000; Handzel et al. 2003), as well as *Plasmodium falciparum* malaria (Githeko et al. 1992), and has traditionally been considered to be an endemic zone for Human African Trypanosomiasis (HAT) (Rutto et al. 2013), although there have been no cases reported for several years (WHO 2013a). The region also has among the highest reported HIV prevalence rates in Kenya of 15.4% in people aged 13 to 34 years (Amornkul et al. 2009).

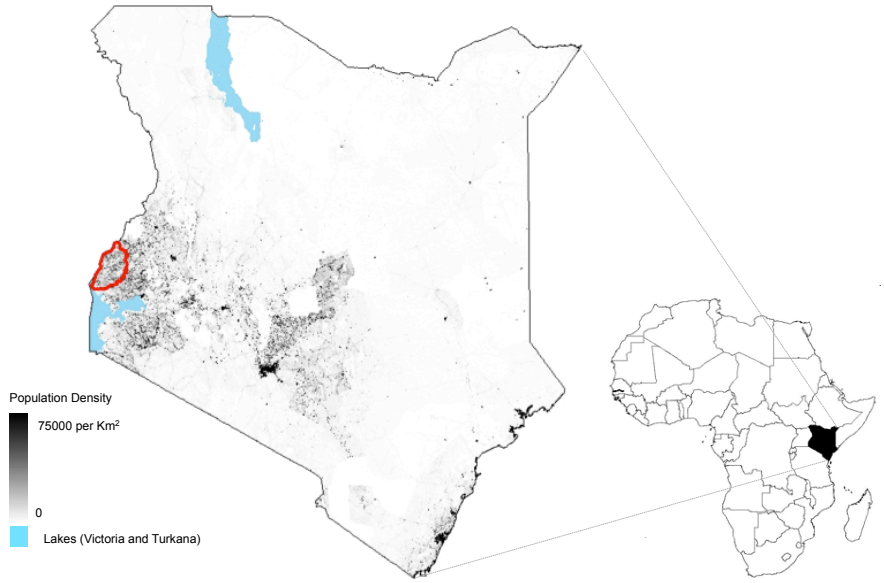


Figure 3.1. The location of the study site in Kenya (highlighted in red) above Lake Victoria.

3.3. Study design

The study design was cross-sectional, in which the primary sampling unit was the household (locally called a homestead), within which all eligible and consenting household members

were recruited. The study sample size was powered primarily to estimate the prevalence of a range of infectious diseases of cattle. An estimated total cattle sample of 2300 were expected to be found in 412 households. Briefly, we considered 5% to be acceptable standard error for the lowest expected prevalence of a bovine disease of 2%, and used a design effect to inflate sample size estimates in order to account for clustering at the homestead level (Fèvre et al *in prep*). The resulting cattle sample size of 2300 head were expected to be found within 412 homesteads, based on local estimates of herd sizes and the frequency of cattle ownership. The human sample size was incidental to that estimated for cattle. Field data collection took place continually between August 2010 and July 2012.

3.3.1. Household selection

Random sampling of homesteads was stratified within sub-locations, the smallest administrative unit in Kenya. The number of homesteads to select per sub-location (between 1 and 8) was proportional to the expected cattle population, so that more homesteads were sampled in sub-locations with larger cattle populations. Cattle population data were available at the division level (the third administrative unit in Kenya) from local Divisional Livestock Production Offices (DPLO) and, in the absence of any other information, were assumed to be proportioned equally in each sub-location falling within each division's boundary.

A number of random points, ranging between 1 and 8, were generated in each sub-location using ArcMap 9.3 (ESRI, Redlands, CA) and the x and y co-ordinates entered into a Garmin eTrex hand-held geographical positioning system (GPS) via the DNR Garmin 5.4.1 extension for ArcView (Minnesota Department for Natural Resources, 2008). The GPS device was used to locate the physical location of each generated random point whilst in the field, and the nearest human habitation within 300 metres was selected for recruitment. In the absence of a household within 300 metres, or following the household head's refusal to participate, a 'back-up' point was randomly generated and recruitment followed in the same manner.

Household recruitment was performed with the assistance of a local chief or village elder, and followed the explanation of study aims and all procedures to the household head. No remuneration was provided for participation, but appropriate management by study clinicians was made available for all readily treatable conditions (including gastrointestinal parasitism and malaria) that were identified during the study.

A household was defined as all people identified by the household head as being occupant at the time of recruitment and typically occupant (to the extent that food is regularly shared from the household pot) within the past 4 weeks. The frequency of household refusals were not recorded, but was very low.

A total of 416 households were recruited, with a total sample size of 2113 people. The average reported household size was 7.6 (range 1 to 30) people (including all age groups), from which our average household sample size was 5.1 (range 1 to 21). Of all eligible individuals present in each household (5 years of age and above and not in the last trimester of pregnancy or any other health issues that might make sampling detrimental, e.g. overt signs of anaemia), we were able to recruit 72.4% (out of 2917).

3.3.2. Data collection within the household

On the sampling day, a detailed questionnaire was performed with the household head and covered data relating to household demography, access to services, known household level risk factors for infectious disease, and durable asset ownership. The questions asked are summarised in Table 3.1. Each household occupant meeting the study inclusion criteria was also interviewed on their education, occupation, food consumption history, contact with livestock and other animals, disease history and current state of health. When participants were less than 12 years of age, a guardian (preferably the mother) was asked to sit with them during questioning to assist with recall. Questionnaires were written in English and translated into the KiSwahili (or KiLuo or KiLuhya) during administration.

A physical exam was performed on all participants by a study clinician (one of two medical officers), who also collected samples for diagnostic testing. A maximum of 25ml of blood venous blood was collected using BD Vacutainer Safety-lok collection set of 21G or 23G from the forearm into plain, heparin, EDTA and Quantiferon-TB Gold tubes. Participants were asked to provide a single faecal sample, collected from the first motion of the day into a collection pot left during recruitment.

All data were recorded on a Personal Digital Assistant (PDA) data entry system (Aceeca MEZ1000 running the 'Pendragon Forms' software) and stored and managed in Microsoft Access databases. A barcode-based system was used to link biological samples to anonymised individuals and homesteads.

The geographic co-ordinates and altitude were collected at a central point within the homestead using a Garmin eTrex hand-held geographical positioning system (GPS).

3.4. Ethical approval

Ethical approval for this study was granted by the Kenya Medical Research Institute (KEMRI) Ethical Review Board (SCC 1701); all participants provided written informed consent.

Table 3.1. Data collected on household level variables. Substantial additional data on livestock management were also collected, but is not shown here. Italicised text in parentheses represents options for multiple option questions.

General	**Animal slaughter**
1. Sex household head	20. Where cattle/pigs slaughtered *(None: Home; Away from home)*
2. Total number of people in household	21. How often cattle/pigs slaughtered
3. Number people per age group	22. Who inspects pigs at slaughter *(None; Myself; Neighbour; Vet; Gov. inspector)*
Water and sanitation	23. Action if *Taenia* cysts found *(Never found; Dispose; Sell; Trim meat; None)*
4. Latrine in compound *(None; Open pit; Partially closed; Closed)*	24. Goats/sheep slaughtered at home
5. Latrine regularly used *[observation]*	25. Household buys meat
6. Latrine accessed by animals *[observation]*	**Medical access**
7. Water source for cooking/drinking in wet/dry season *(Borehole; Dam/Pond; River; Well; Spring; Piped; Roof capture; Other)*	26. Medical facilities used by majority of household *(Community health worker; Traditional; Church; Chemist; Hospital; Health centre; None; Other)*
8. Water treatment type used in the last month *(None; Boil; Chlorine; Iodine; Filter; Other)*	27. Distance to most used medical facility
9. Village flooding in last 12 months	28. Transport to most used medical facility *(Walk; [Own] Bicycle; [Own] Motorbike; [Own] Car; Matatu; Ambulance; Other)*
10. Village drought in last 12 months	29. Cost of transport to most used medical facility
Household assets	**Livestock**
11. Main household cooking fuel *(Firewood; Charcoal; Gas; Paraffin; Solar; Electric)*	30. Number of cattle per sex and age group
12. Number dwellings occupied	31. Number of pigs per sex and age group
13. Number dwellings with roof/thatch/tiles/other roof *[observation]*	32. Number sheep
14. Number dwellings with mud/unburnt brick/burnt mud brick/brick and cement/cement/timber/stone/ other walls *[observation]*	33. Number goats
15. Number of dwellings with earth/cement/tiled/wooden/other floor *[observation]*	34. Household keep poultry
16. Electricity in household *(Mains; Generator; Car battery; Solar power; None)*	35. Household keep small animals
17. Household electric goods *(Radio; Mobile phone; Phone charger; TV; None)*	36. Household grows crops
18. Household furniture *(Cupboard; Wooden bed; Bed net; Sofa; Clock; Watch; Sewing machine; Torch; None)*	37. Reason for growing crops *(Home consumption, Cash)*
19. Household transport *(Bicycle; Motorbike; Car; None)*	

3.5. Laboratory processing

Blood and faecal samples were tested for a very wide range of zoonotic and non-zoonotic infectious agents that were expected to be endemic in the study area. These agents, and the diagnostic tests used, are summarised in Table 3.2.

A flow chart showing the steps in the processing of human samples within the field laboratory is given in Figure 3.2.

Faecal samples were prepared using standard protocols for the Kato-Katz and Formal Ether techniques (Katz et al. 1972; Allen & Ridley 1970) and examined under light microscopy. The presence or absence helminth and protozoal gastrointestinal parasites was recorded and a quantitative estimate of the number of eggs per gram (EPG) of faeces calculated for each parasite identified. Samples were additionally prepared using Ziehl-Nielsen staining to enable the identification of *Cryptosporidium* species. Remaining material was stored in 5% saline with 0.3% Tween-20 at room temperature for later analysis by copro-antigen tests (for *Taenia* spp.).

Thin and thick blood smears were made in the field using blood remaining in the butterfly apparatus following venous sampling. Slides were stained using 10% Giemsa and examined under 100x with oil immersion objective lens. Haemoparasites observed were recorded qualitatively (present/absent) and semi-quantitatively on the basis of a standard intensity scale: 1 – 10 parasites per high power field = +; 11-100 = ++; 1 – 10 in every field = +++; more than 10 parasites in every field = ++++.

The buffy coat and the red blood cell/buffy coat interface from centrifuged haematocrit tubes containing heparinised blood were examined under 100x oil immersion and at the 10x power for the presence of *Trypanosomes* and *Ricketsiae* (the Haematocrit centrifugation technique, or the "Woo Method") (Woo 1971). The presence of motile organisms was further evaluated by transferring the buffy coat to a microscope slide and examining 100 microscope fields at x10 power.

Haemaglobin concentrations were measured using Hemocue system on heparinised blood, and individuals were described as anaemic using the age and sex cut-offs defined by the WHO (2011), after adjustment for altitude (which ranges between 1000 to 1600 metres above sea level in the study area). In order to assess micronutrient status, serum concentrations of ferritin, retinol binding protein (RBP), as well as acute phase proteins (CRP and AGP) were measured using a sandwich ELISA (Erhardt et al. 2004) at an external laboratory (VitMin Lab, Willstaett, Germany). Iron deficiency was defined as a ferritin

concentration < 15 μg/L (since all individuals were greater than 5 years of age), after adjustment for the effect of sub-clinical inflammation using acute phase proteins (CRP and AGP) and the 4-group approach described by Thurnham et al. (2010). Vitamin A deficiency was defined as RBP < 0.7 μmol/L, after adjustment for inflammation using both CRP and AGP and the 4-group approach described by Thurnham et al. (2003).

Blood collected in serum tubes was spun at 3000rpm for 20 minutes, and aliquoted into 2ml barcoded cryovials before serological testing, or storage at -40°C for later analysis.

Table 3.2. Summary of the faecal and blood-based tests used to identify infectious agents in people.

Infectious agent (disease)	Sample	Test	Supplier
Bacteria			
Brucella abortus/*B. melitensis* (brucellosis)	Serum	Rapid immuno-chromatographic flow assay (IgG, IgM)	Life Assay, South Africa
Coxiella burnetii (Q-fever)	Serum	ELISA (IgG)	Serion-Virion GmbH, Germany
Mycobacterium tuberculosis (TB)	Heparinised whole blood	QuantiFERON-TB test (gamma interferon)	Cellestis Limited, Australia
Viruses			
HIV	Heparinised whole blood	SD Bioline HIV 1/2 3.0 (rapid strip test)	Standard Diagnostics Inc., South Korea
Rift Valley Fever	Serum	cELISA	BDSL diagnostics, South Africa
Protozoa			
Plasmodium falciparum (malaria)	Blood smear (whole blood)	Light microscopy (thick and thin)	-
Trypanosoma brucei (sleeping sickness)	Heparinised whole blood	Light microscopy; Woo test (see text)	-
Cryptosporidium spp. (cryptosporidiosis)	Faeces	Ziehl-Neelsen staining	-
Giardia intestinalis (giardiasis)	Faeces	Formal ether concentration (FEC)	-
Entamoeba histolytica (ameobiasis)	Faeces	FEC	-
*Blastocystis hominis**	Faeces	FEC	-
Balantidium coli (balantidiasis)	Faeces	FEC	-
*Endolimax nana**	Faeces	FEC	-
Helminths			
Wucheria bancroti (filariasis)	Blood smear (whole blood)	Light microscopy	-
Ascaris lumbricoides (ascariasis)	Faeces	Kato-Katz; FEC	Kato-Katz equipment from Simon Brooker
Ancylostoma duodenale/Necator americanus (hookworm)	Faeces	Kato-Katz; FEC	-
Trichuris trichiura (trichuriasis)	Faeces	Kato-Katz; FEC	-
Strongyloides stercoralis (strongyloidiasis)	Faeces	FEC	-
Schistosoma mansoni (schistosomiasis)	Faeces	Kato-Katz; FEC	-
Fasciola spp. (fascioliasis)	Faeces	Kato-Katz	-
Taenia solium/T. saginata (taeniasis)	Faeces	FEC; Copro-Ag ELISA	Pierre Dorny & Sarah Gabriel
Taenia saginata (cysticercosis)	Serum	HP10-Ag ELISA	Leslie Harrison & Micheal Parkhouse
Hymenolepsis spp. (hymenolepiasis)	Faeces	Kato-Katz	-

* Pathogenicity is uncertain

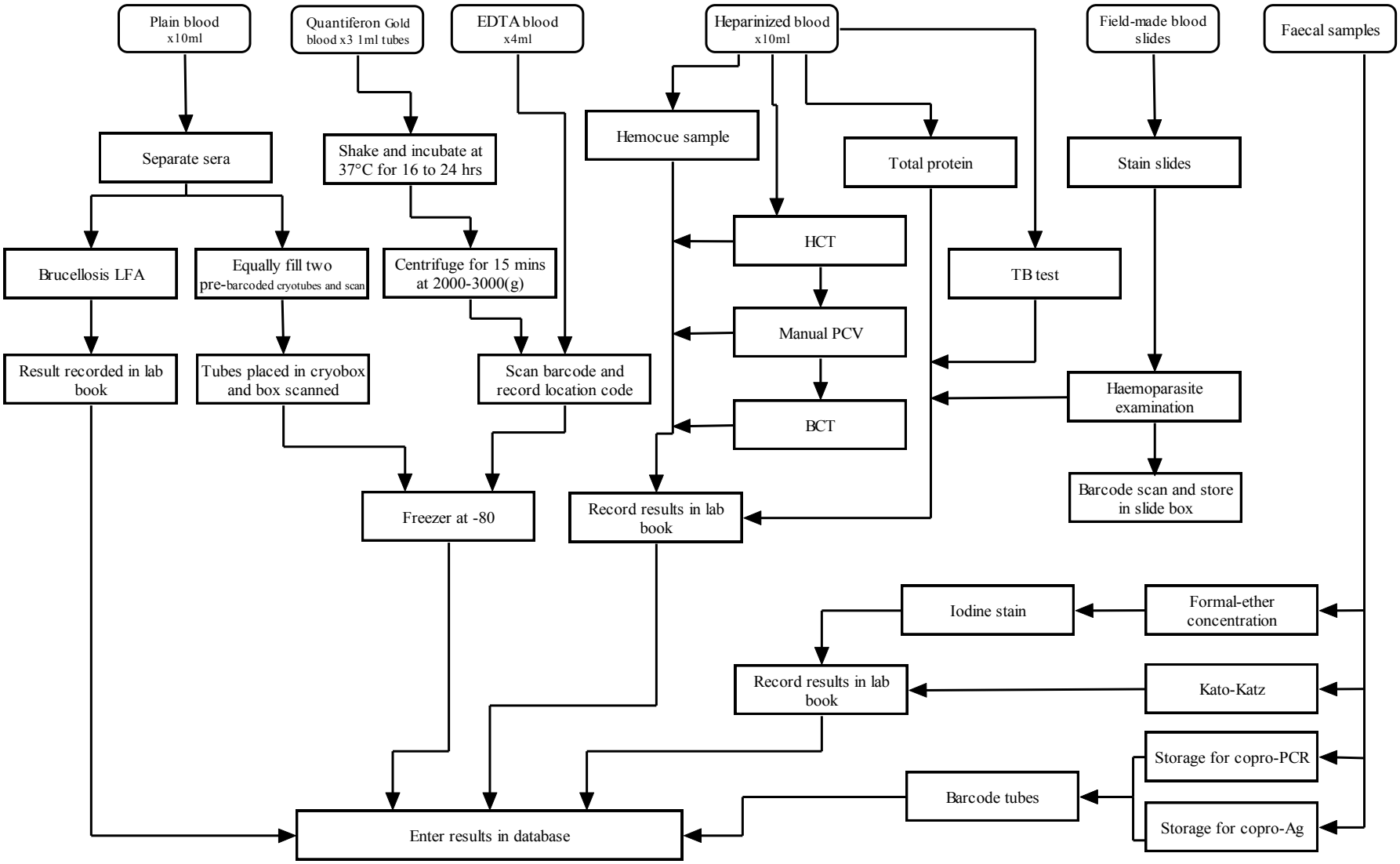


Figure 3.2 Flow chart for the field-laboratory based processing of samples collected from people

Chapter 4

Infection in context: exploring infectious disease risk in a rural community in western Kenya

4.1. Introduction

People living in rural areas in sub-Saharan Africa may be at high risk of infection with a wide range of infectious agents (Petney & Andrews 1998; Buck et al. 1978; WHO 2012b). The burden of preventable and treatable communicable disease within these communities perpetuates poverty (Fox et al. 2004; Audibert 2010), reduces individual well-being and quality of life (Deribew et al. 2009; Fürst et al. 2012), and is an important cause of mortality (Mathers & Loncar 2006).

In populations in which multiple infectious diseases are endemic, an individual's risk of infection with a particular pathogen is likely to depend on a complex interplay of factors that relate to their exposure and susceptibility, and which may be conditioned by the social, cultural, political, economic and environmental contexts in which they live (Diez-Roux 1998; Diez-Roux & Aiello 2005; Allotey et al. 2008). Factors operating at the group level (e.g. the household, the village) may influence outcomes independently of individual characteristics, or modify how these are related to health outcomes (Diez-Roux 1998). Such contextual phenomena will tend to result in health outcomes clustering within groups (Merlo, Yang, et al. 2005c). Individuals living in the same household (Pullan, Kabatereine, et al. 2010b; Brooker, Alexander, et al. 2006a; Walker et al. 2011), or in the same village (Ashford et al. 1993), may therefore be more similar in their risk for a particular outcome than the general population. Contextual effects can be particularly profound for infectious disease (Morgenstern 1995), not least since infected members of a group often infect other members of the same group, directly, via vectors, or via the environment (Cairncross et al. 1996). Indeed, much of an individual's risk of infection may be determined by the infection or immune state of the people around them (Fine et al. 2011). It is well known, for example, that the use of sanitation as an intervention to reduce environmentally-transmitted infection, such as that due to the soil-transmitted helminths, is most successful when whole communities are targeted (Esrey et al. 1991; Appleton et al. 2009; Asaolu & Ofoezie 2003),

and that individuals living in households with good standards of sanitation may still be at high risk of gastrointestinal parasitism if their neighbours do not also have good standards (Chongsuvivatwong et al. 1996; Feachem et al. 1983). Understanding individual risks can generally be improved by examining effects operating at multiple levels (Schwartz 1994; Susser & Susser 1996).

Spatial analysis can provide a powerful means with which to identify spatially heterogeneous contextual effects. Such effects might explain why disease risk varies in people with the same individual characteristics, but living in different social or biophysical environments (Petronis & Anthony 2003; Chaix et al. 2005; Logan 2012). Alternatively, clustering of adverse health outcomes may occur as a result of aggregations of individual risk factors within certain regions, or compositional effects (Curtis & Rees Jones 1998). Exploratory spatial analysis can therefore provide a basis for further hypothesis-driven research that seeks to disentangle some of these contextual and compositional effects on individual disease risk (using spatial or non-spatial methods (Soares Magalhães, Biritwum, et al. 2011b; Pullan et al. 2012)).

In this chapter, we report on the individual- and household-level prevalence of infection with, or exposure to, a very wide range of infectious agents in a rural community in western Kenya. Prevalence provides a summary of individual distributions (Rothman & Greenland 1998), and could be considered to represent the average probability of an outcome being present in a randomly selected member of the population. Individual parameters aggregated at the group level tend to do a poor job of predicting outcomes when applied back to individuals (rather than the population of individuals), the so-called "tyranny of the average" (Merlo & Wagner 2013), and, as we have described, an individual's risk of infection is likely to be highly specific to them and the context in which they live. To assess the extent to which residential location may be important (Booth 2006), we examined how an individual's risk of infection varies in space. Of particular interest was the identification of spatial congruence (or 'co-incident overlap' (Lello et al. 2013)) in risk for multiple pathogens that might indicate the existence of shared contextual or compositional effects, or provide evidence for potentially important biological interactions between pathogens. To further explore contextual effects operating at the household level, we also quantified the extent to which infection risk varies between individuals in different households, and between individuals in households in different administrative divisions.

4.2. Methods

All data for this chapter were collected as part of the 'People, Animals and their Zoonoses' (PAZ) project. Background on the study design is given in chapter 3.

4.2.1. Estimating individual-level infection prevalence

Our complex survey design involved sampling individuals living in randomly selected households in sublocations. Members of the same household are likely to be correlated in their disease status (due to a range of contextual and compositional effects that they share). This within-cluster (or intra-class) correlation should be accounted for in order to derive population summaries with appropriate standard errors (Lumley 2004). Furthermore, and since sample sizes and population numbers varied between sublocations, sampling weights (the inverse probability of selection) were required to return our distorted sampling approach to one that approximates simple random sampling (Lumley 2004).

We used design-based inference to adjust individual infection prevalence estimates and their standard errors on the basis of this complex design (Lumley 2010). There were a large number of 'singleton' primary sampling units (PSUs, our households) at the sublocation level (i.e. sublocations in which a single household was sampled (n = 28)), hence sublocations (n = 288) were aggregated by division (n = 17), which was used as a stratifying variable. A unique identifier was used for each household to account for clustering.

Sampling weights were calculated as $1/\pi_1$, where π_1 is the sampling probability for each individual in each division, estimated as the fraction of the number of individuals sampled and the total number of people per division. The total population size per division was derived from the 2009 census (Open Kenya 2014). Sampling with replacement was assumed (Lumley 2004).

Design-based adjustment was implemented using the *svydesign* procedure in the *Survey* package in R (Lumley 2004). We examined the relationship between the prevalence of each infection and gender using the Wald statistic, as proposed by Koch et al. (1975) for complex survey designs.

4.2.2. Estimating household-level prevalence

Group 'infection' status can be classified on the basis of tests applied to individuals, and a threshold of *n* or more positive cases used to define presence or absence of infection within

the group (Christensen & Gardner 2000; Hanson et al. 2003). We considered the presence of at least one infected person to represent household infection for all infectious agents under study. However, it is necessary to account for test error at this threshold, and for the fact that the probability of misclassification may change on the basis of sampling effort per group. These issues have been extensively characterised and discussed in the veterinary literature, where inference on farm-level infection is a common goal (Branscum et al. 2004; McV Messam et al. 2008). In that field, the gold standard approach to estimate herd prevalence is through the use of binomial probability sampling within a Bayesian framework, which allows the derivation of both the distribution of within-herd prevalence as well as the proportion of infected herds, with the use of mixture distributions to allow for zero infection prevalence, and the incorporation of prior distributions describing diagnostic test performance (diagnostic sensitivity (Se) and specificity (Sp)) (Lewis & Torgerson 2012; Branscum et al. 2004). In the application of these models, counts for test positive individuals (r_i) out n sampled individuals in the *i*th herd (group) are distributed as: $r_i \sim binomial(n_i, Se \times P_i + (1 - Sp) \times (1 - P_i))$, where P_i is the within herd prevalence, modelled as a zero inflated distribution (i.e. describing exposed or unexposed) with probability, τ (Branscum et al. 2004).

Our use of these group-prevalence models was somewhat limited by the inherently small number of people in each household: when the sample size per group is large relative to the group size, the binomial sampling approximation described above is no longer appropriate (Su et al. 2004). In such cases, it is suggested that a hypergeometric approach based on finite probability sampling without replacement should be used (Branscum et al. 2004). Group prevalence models relying on hypergeometric sampling are considerably more complex than those that can make a binomial sampling approximation (Hanson et al. 2003), and we could find no examples of these being fit in BUGS-based (or other) software. However, hypergeometric approaches to estimate the prevalence of infection in a *single* group are more straightforward (Branscum et al. 2004), and several stand-alone packages have been developed that use hypergeometric sampling to calculate single-group prevalence with adjustment for test error. The Bayesian Disease Freedom (BDFree) software[2], for example, uses a hypergeometric approach to estimate within-group prevalence (P_T) on the basis of the number of individuals sampled (n) and the number of positive cases detected (r) (Johnson et al 2004). The software, being Bayesian in its inference, allows the inclusion of prior information on test performance as well as expected within-group prevalence. We used

[2] http://www.epi.ucdavis.edu/diagnostictests/bayesfreecalc.html

BDFree to estimate the probability that the P_T for each infection in each household (and group of people therein) was at or lower than a pre-specified cut-off, x (distributed as $x|(P_T, Se, Sp) \sim$ hypergeometric(n, r, P_A), where P_A is the observed within-group prevalence). We used a very low cut-off value of x = 0.0001 ≈ 0, as suggested by McV Messam et al. (2008), and the posterior probability of household infection (P_{INF}) was then 1- Pr($P_T \leq x$) (or 1 minus the probability that the within-household prevalence was zero).

This procedure was repeated for each infection in each household, and the resulting estimates of P_{INF} averaged to derive an estimate that can be considered broadly equivalent to the between-household prevalence (or the probability that a randomly selected household contains at least one person with an infection of interest) with adjustment for test error. It is important to note that this approach takes no account of the prior probability that a household is infected (i.e. there is no zero inflation by probability τ, described above). As such, there may be a tendency for household-prevalence estimates for rare outcomes to be biased upwards (McV Messam et al. 2008).

Estimates of diagnostic sensitivity and specificity for each test for each infection (and the resulting hyperparameters for the Beta distributions) were derived from the literature. Within-household prevalence was modelled using an uninformative prior (i.e. Beta(1,1)).

To enable comparison of these adjusted estimates with the observed data, we also derived between-household prevalence as the proportion of "infected" households defined on the basis of a standard cut-off of one infected person per household (i.e. without adjustment for sampling effort or diagnostic error). Point estimates and standard errors were adjusted using the *svydesign* procedure described in section 4.2.1, with sampling weights calculated as the fraction of the total number of households in the division and the number of households sampled. Household counts per division were derived from the 2009 census (Open Kenya 2014). No clustering variable was used.

4.2.3. Spatial analysis

We explored the spatial distribution of the pathogens under study using household infection status.

Kernel density estimation of the distribution of household risk

Kernel density estimation (KDE) provides a means of deriving a continuous spatial surface that describes the density/intensity of a set of unmarked spatial points. The approach involves placement of a regular grid over a region and defining a (typically Gaussian) kernel with a specified radius (bandwidth) around each point. This allows the estimation of a continuous distance-decay density over the whole area under study (Pfeiffer et al. 2008). This approach can be applied to cross-sectional data to derive a spatially continuous risk surface as the ratio of the density (or intensity when estimates are per unit area) of case events and the density of all sampled events (Bithell 1990; Sarojinie Fernando & Hazelton 2014). This approach has been used widely in veterinary epidemiology (Pfeiffer et al. 2008), and is also increasingly being used for human infectious disease, particularly for vector-borne infections (Mosha et al. 2014; Medeiros et al. 2012; Simarro et al. 2012), as well as for HIV (Tanser et al. 2009). Analysis using kernel smoothed surfaces has also already been applied to the PAZ dataset in order to explore the overlap in the spatial distribution of taeniasis and cysticercosis in people and pigs (Wardrop et al *in prep*).

To date, there appears to have been little work (in the veterinary or medical literature) to identify the uncertainty in smoothed KDE surfaces that may occur as a result of misclassification of the point process (in our case, the definition of household-level infection). We attempted to incorporate this uncertainty by using the probability of infection (P_{INF}) derived from the Bayesian hypergeometric sampling procedure (described in section 4.2.2.). Household infection status (0 or 1) was classified over a large number (3000) of Bernoulli trials, with P_{INF} defining probability of success for each trial. At each step (i.e. for each trial across all households), we estimated the kernel intensity of case (positive) households per unit area using a fixed bandwidth of 5000 metres. Kernel intensity surfaces were derived using the *sparr* package (Davies et al. 2011) in R, which allows account for edge effects (due to incomplete data that occurs at the boundary of the study area).

Each of the resulting smoothed case intensity surfaces was divided by a single smoothed surface representing the intensity of all households in the study area (i.e. there were 3000 possible numerators and a single denominator), estimated using the same bandwidth as that for case intensity. The resulting smoothed surfaces could be considered to represent the distribution of smoothed surfaces of household risk for each infection across the study area, with adjustment for the effects of diagnostic test error and sampling effort on the classification of household infection status. We extracted, and present the average, 5^{th} and

95th percentile from this distribution for each infection. This procedure is summarised in Figure 4.1, and the R code used to automate it is given in Appendix 4.3.

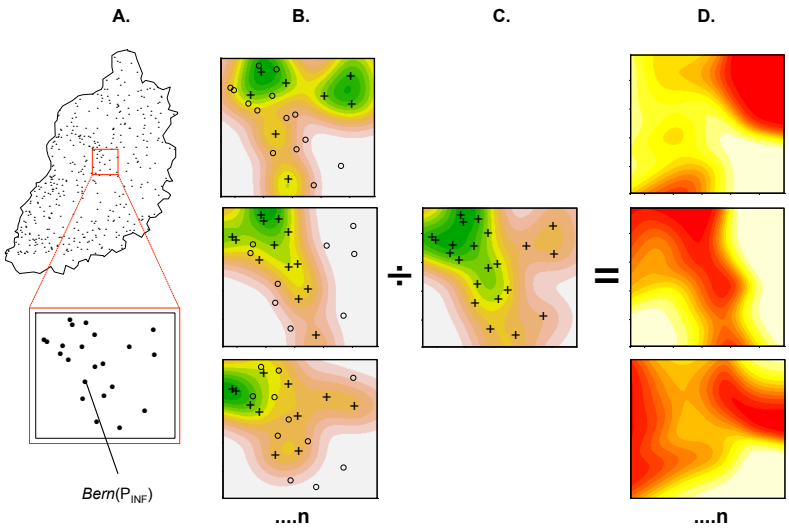


Figure 4.1. Summary of the KDE procedure to derive the distribution of household risk estimates. A: Infection status of each household is defined over multiple iterations (n = 3000) on the basis of a single Bernoulli trial; B: For each iteration, kernel intensity of case (positive, indicated by +) households is derived; C: Each case intensity surface is divided by a single measure of all household intensity; D: Estimate of household risk derived for each iteration (with redder colours indicating increased risk in above example), producing a distribution of smoothed surfaces with account for misclassification.

Only data for those infections in which the observed household prevalence was greater than 2.5% (i.e. >10 positive households) were used. Bandwidth choice for KDE is much-debated, and can be derived using range of data-driven approaches or on the basis of the domain knowledge of the user (Pfeiffer et al. 2008). We used a uniform value of 5000 metres for each infection, which was broadly expected to represent the distance over which variation in disease risk could be expected to occur.

In order to compare the outputs from this procedure with the observed data (i.e. without adjustment for test misclassification), we also defined cases as households in which at least one person with each infection was identified. A single case intensity surface was derived for each infection which was divided by the intensity of households at risk (resulting in a single layer describing risk without adjustment for misclassification).

To quantitatively explore spatial relationships in household risk, Pearson's product moment correlation was estimated for the comparison between the average (simulated) KDE risk estimates for each pair of pathogens. Autocorrelation present within spatial data produces well known biases in the estimation of correlation coefficients and variances (Dale & Fortin 2009). We therefore used the modified t-test approach described by Clifford et al. (1989), which allows the assessment of the presence of autocorrelation within spatial data and reduces the effective sample size accordingly. This modified t-test procedure was implemented using the *SpatialPack* package (Osorio & Vallejos 2014) in R.

Using KDE estimates to derive an estimate of the risk of household polyparasitism

The smoothed surfaces described above can be considered to estimate spatial variation in household risk, or the household prevalence. The value of each pixel of the raster layer making up the smoothed risk surfaces for each infection can therefore be considered to provide an estimate of the probability that a randomly selected household within each pixel has at least one person with the infection of interest. It is possible to combine these unique probabilities to derive an estimate of the probability that the same randomly selected household has *at least n* infections out of the *m* possible infections (i.e. out of all infections for which we have a spatial risk surface).

Consider the scenario in which we have risk surfaces for four infections A, B, C and D, with associated estimates of risk, Pa, Pb, Pc and Pd. For a particular point on our map, we would like to estimate the probability that a randomly selected household has at least one of the infections of interest (i.e. at least one infection is present in at least one person). This would simply be 1-(1-Pa*1-Pb*1-Pc*1-Pd), or one minus the probability of obtaining zero events. If we wanted to estimate the probability of *at least two* infections being present in the household, we would first need to know the probability for each of the 6 mutually exclusive possible combinations (AB, AC, AD, BC, BD, CD). For AB this would be Pa*Pb*1-Pc*1-Pd; for AC it would be Pa*1-Pb*Pc*1-Pd, and so on. Summing these probabilities would give us the probability of two diseases being present. To derive the probability of *at least* two events, we would need to combine this with the probability of three disease being present ((Pa*Pb*Pc*1-Pd) + (Pa*Pb*1-Pc*Pd) + (Pa*Pb*Pc*1-Pd) + (1-Pa*Pb*Pc*Pd)) and that for four diseases (Pa*Pb*Pc*Pd). If Pa, Pb, Pc and Pd were all 0.5, the probability of at least two diseases being present in our randomly selected household would be around 0.7.

This general approach was extended for every pixel of a continuous raster across the whole surface of our study area. The probability of household infection with each infectious agent

(m) for each cell was based on the average probability from the KDE procedure (described above, with adjustment for test misclassification). The R code for implementation is given in Appendix 4.3.

Identifying spatial clustering of household infection

A commonly used approach to detect clustering within spatial point data is the spatial scan statistic (Kulldorff 1997). This technique involves the construction of a series of circles of increasing size across the spatial (and potentially temporal) dimensions of a study area and comparison of the risk inside the 'cluster' with that outside. The likelihood function for the probability model used (which can be Bernoulli, Normal, Poisson etc) is maximised over all clusters and significance tests are performed through Monte Carlo hypothesis testing by comparing the rank of the maximum likelihood from the observed data with the maximum likelihoods from a set of random data permutations. This approach has been widely applied for spatial cluster detection in infectious disease research (Pullan et al. 2012; Auchincloss et al. 2012).

The spatial scan statistic was implemented in SatScan version 9.0 (www.satscan.org). To identify spatial clusters for each infection, we used the observed (unadjusted) data of household infection (i.e. at least one infected person) within a Bernoulli model, 999 iterations (allowing estimation of p-values down to 0.001) and a cluster size up to a maximum of 20% of observations (to avoid spurious clusters that include a large proportion of the total population).

To explore potential clustering in species richness (i.e. the total count of unique pathogen species within a household), we also used a discrete Poisson model for the observed (i.e. unadjusted) count of all species identified, together with the number of people sampled. Under the null hypothesis, the expected number of events (unique species) is proportional to the population size at each location (Kulldorff 1997). Whilst the infectious agents under study (shown in Table 3.2 in chapter 3) have a range of transmission routes, infection with the gastro-intestinal nematode and protozoal parasites (excluding *Taenia* spp. *Fasciola* spp., *Balantidium coli*, and *Schistosoma mansoni*) can broadly be considered to follow similar routes of transmission (i.e. exposure to faecal contamination or faecally contaminated food, soil or water). To test whether this large group of broadly similar parasites might be driving any observed clustering in species richness, we repeated the Poisson spatial scan statistic on a count of these parasites alone and the count of all infectious agents with the gastrointestinal

parasites excluded. In all cases, the model used 999 iterations and a cluster size up to a maximum of 20% of observations.

A potential criticism of the spatial scan statistic is its reliance on a circular scanning window, and therefore its inability to detect irregularly shaped clusters. The spatial *relative* risk (the parameter being approximated by the Bernoulli spatial scan statistic) can also be estimated using a kernel smoothing approach as the ratio of the density of cases and controls (Bithell 1990; Sarojinie Fernando & Hazelton 2014). Moreover, asymptotic methods are now available to calculate the probability that the observed relative risk is higher than the null hypothesis of unit relative risk for the whole surface being mapped (Hazelton & Davies 2009). This allows p-values to be plotted over the smoothed surface, and therefore the identification of areas of significantly elevated relative risk (or 'hot spots') which are not constrained by cluster shape.

We used the *sparr* package in R (Davies et al. 2011) to derive the ratio of case density (households with at least one infection) and control density (households without any infection) and to estimate and plot asymptotic p-values over the resulting relative risk surface for each infection.

For all tests of spatial clustering, we used the observed data only (i.e. without adjustment for test error).

4.2.4. Multilevel analysis to partition variance

Multilevel analysis is now widely used in epidemiology to simultaneously investigate the effects of factors at multiple levels (Diez-Roux & Aiello 2005; Blakely & Woodward 2000). These levels can be defined in many ways, but in our case, individuals (n=2113) were naturally grouped within households (n=416), and households grouped within administrative divisions (n = 17). These groupings represent the nesting of sources of variability in the observed data at each of the defined levels (Diez-Roux & Aiello 2005). In the many applications of these hierarchical models in infectious disease epidemiology, this variability is treated only as statistical 'noise' to be controlled in order to derive unbiased estimates and appropriate precision for standard errors (Snijders & Bosker 2011). Such control is perfectly appropriate, but the presence of inter-group variation at each of these levels (i.e. the clustering of observations) is rarely considered as yielding important information by itself, and measures of variation (such as the intra-class correlation co-efficient (ICC)) are rarely reported. Several authors have suggested a greater focus on measures of variance and

clustering as potentially highlighting the existence of effects that operate at the level of the group, or contextual effects (Petronis & Anthony 2003; Merlo 2003; Merlo et al 2005a; Merlo et al 2005b; Merlo et al. 2006; Larsen & Merlo 2005). To explore the possible existence of such effects, particularly in the context of multiple infections, and to provide a basis for the subsequent investigation of household-level factors that might explain the observed variation, we derived a series of simple multilevel regression models and focused on the variation present in individuals between group (*sensu* Merlo et al 2005a).

Model specification

For each infection, we derived a three-level logistic regression model with household and administrative division as random effects. These models estimate the probability of infection together with the variance at the intercept for household ($\sigma^2{}_H$) and division levels ($\sigma^2{}_D$). Whilst our primary intention was to identify general contextual effects operating at the group level, age and/or gender are thought to be important for virtually all of the diseases under study. We therefore included age with 5 categories (5-9 years; 10-14; 15-24; 25-39; 40+) together with gender as fixed effects at the individual (first) level to account for these potentially important compositional effects. Only infections with an (observed) individual level prevalence of greater than 5% were examined.

In addition, we used a two-level Poisson regression model to examine species richness (i.e. the count of all observed species) at the household level with division at the second level as a random effect (and therefore estimated a single group variance at the division level, $\sigma^2{}_D$). The two-level Poisson model did not include any fixed effects, but did incorporate a log transformed count of the number of individuals sampled in each household as an offset. This provides some account for the variable sampling effort per household, enabling the outcome to be modelled as a rate rather than a count (Dohoo et al. 2009).

Quantifying general contextual effects

The variance partition co-efficient

We calculated the variance partition co-efficient (VPC) for the outputs from each of the logistic regression models for each infection using the latent variable method (Snijders & Bosker 2011). This approach converts the (unobserved) individual-level variance from the probability scale to the logistic scale so that it is comparable with the group level variance (Merlo et al. 2006). We therefore assume that risk of infection is a latent variable with some

threshold, beyond which an individual becomes infected. The variance of this underlying individual variable follows a logistic distribution, equal to $\pi^2/3$ (i.e 3.29) (Snijders & Bosker 2011), so that VPC at the division ($_D$) and household ($_H$) level can be calculated as:

$$VPC_D = \frac{{\sigma^2}_D}{{\sigma^2}_D + {\sigma^2}_H + \pi^2/3}$$

$$VPC_H = \frac{{\sigma^2}_D + {\sigma^2}_H}{{\sigma^2}_D + {\sigma^2}_H + \pi^2/3}$$

The VPC represents the correlation in the probability of infection between two individuals randomly selected from the same division (VPC_D) or from the same household in the same division (VPC_H), and therefore provides a measure of within-group clustering that, in its general form, is equivalent to the intra-class correlation co-efficient (ICC) (Merlo et al. 2012).

There is no Poisson regression analogue of the latent variable method to estimate VPC/ICC, but exact methods have been shown to be suitable for estimating correlation between two observations in the same group (Stryhn et al. 2006). The exact methods estimate variance at each level of a two-level model as (Dohoo et al. 2009):

$$\text{Level 1:} \quad \sigma(1) = \exp\left(\beta X + \frac{{\sigma^2}_D}{2}\right)$$

$$\text{Level 2:} \quad \sigma(2) = \exp(2\beta X + 2{\sigma^2}_D) - \exp(2\beta X + {\sigma^2}_D)$$

So that:

$$\text{ICC} = \frac{\sigma(2)}{\sigma(1) + \sigma(1)}$$

Where *BX* is are fixed effects at the individual level (in our case the logarithmic offset).

Median odds ratio

The median odds ratio (MOR) (Larsen & Merlo 2005) provides a measure of the heterogeneity in variation between groups (Merlo et al. 2012), and is the median value of the odds ratio between a group at higher risk and a group at lower risk when randomly picking two groups. It can be conceptualised as the increased risk of infection that (in median) an individual would have if moving to another household or division with higher risk, and provides a means with which to investigate the existence of contextual effects for simple and complicated models alike (Merlo et al. 2006). It can be calculated as the median value of the

distribution describing comparisons between each pair of individuals in which the individual with the higher odds is the numerator (so the ratio is always 1 or more) (Larsen & Merlo 2005). This generalises to (Merlo et al. 2012):

$$\mathrm{MOR}_D \approx \exp\left(0.95 * \sqrt{{\sigma^2}_D}\right)$$

$$\mathrm{MOR}_H \approx \exp\left(0.95 * \sqrt{{\sigma^2}_D + {\sigma^2}_H}\right)$$

The concept of MOR appears not yet to have been embraced by the infectious disease community, but MORs were recently reported in a multi-level analysis of risk factors for malaria in Ethiopia (Woyessa et al. 2013). The MOR is generalisable to multi-level Poisson regression models (Larsen & Merlo 2005), and equivalent median rate ratios (MRR) have been reported for multi-level models from other fields (Ervasti et al. 2012).

4.3. Results

4.3.1. Individual infection

The individual-level prevalence for those infectious agents in which at least one infected person was identified are presented in Table 4.1. We did not observe (but considered possible) infection with *Isospora spp.*, *Cyclospora spp.*, *Dientamoeba fraglis*, *Trichostrongylus spp.*, *Schistosoma bovis*, *Wucheria bancroti* or *Trypanosoma brucei*. There was evidence for a relationship with gender for several infections, with significantly higher prevalence in males for *Strongyloides stercoralis*, *Schistosoma mansoni*, *Plasmodium falciparum*, and hookworm (Table 4.1.). Females were at significantly higher risk for cysticercosis (i.e. the presence of a *Taenia solium* cyst, on the basis of a HP10 ELISA), but there was no evidence of a relationship between sex and taeniasis (the presence of a *T. solium* or *T. saginata* worm in the small intestine, on the basis of Copro-ELISA). Females were also at significantly higher risk for infection with HIV, *Entamoeba histolytica/dispar*, and *Trichuris trichiura*. Age prevalence profiles for the common infections are given in Appendix 4.1.

Owing to incomplete faecal or blood collection from some individuals, estimates of polyparisitism (where each infection under study had the potential to be observed) were available for 1728 out of 2113 (82%) individuals and 379 out of 416 (91%) households. The average parasite species count from these people (out of a total of 21 possible infections

(listed in Table 4.1), using combined data from Copro-ELISA and HP10-ELISA to represent all *Taenia* infection) was 1.91 (95% CI 1.87 – 1.94), with a range of 0 to 7, and a median of 2. Two hundred and fifty nine individuals (14.9%) had an infection count of zero while 204 (11.8%) had 4 parasites or more, 59 (3.4%) had 5 or more, 14 (0.8%) had 6 or more, and one individual had 7 concurrent infections.

There was no difference in parasite species count between males and females ($t = 0.2$, $p = 0.82$), but male children tended to have a higher count than female children, while the count was higher in adult females than males, although 95% confidence intervals overlapped for each age group (Figure 4.2).

Table 4.1. Individual and gender-stratified prevalence estimates for the gastrointestinal parasites, bacterial and viral infections under study

Infection	Adjusted[1] prevalence (95% CI)	Male (%)	Female (%)	p value[2]
Gastrointestinal parasites				
Balantidium coli	0.02 (0 – 0. 1)	-	-	-
Fasciola spp.	0.04 (0 – 0. 1)	-	-	-
Entamoeba hartmanni	0.1 (0 – 0.2)	-	-	-
Endolimax nana	0.1 (0 – 0.2)	-	-	-
Hymenolepis spp.	0.2 (0 – 0.3)	-	-	-
Taenia spp. (eggs)	0.3 (0 – 0.5)	-	-	-
Blastocystis hominis	0.6 (0.1 – 1.1)	-	-	-
Cryptosporidium spp.	0.6 (0.2 – 1.0)	-	-	-
Strongyloides stercoralis	2.9 (2.1 – 3.8)	3.9	2.1	0.023
Giardia spp.	3.2 (2.3 – 4.0)	4.0	2.5	0.09
Taenia solium (HP10-ELISA)	5.8 (4.4 – 7.2)	4.6	7.0	0.04
Schistosoma mansoni	5.9 (3.7 – 8.1)	7.2	4.8	0.009
Trichuris trichiura	10.0 (8.2 – 11.7)	7.6	12.0	0.002
Ascaris lumbricoides	10.4 (8.1 – 12.7)	9.7	11.1	0.33
Iodamoeba butschlii	14.2 (12.4 – 16.0)	13.4	15.0	0.42
Taenia spp. (Copro-ELISA)	19.7 (16.7 – 22.7)	20.7	18.8	0.29
Plasmodium falciparum	29.4 (26.8 – 32.0	32.1	27.0	0.02
Entamoeba histolytica/dispar	30.1 (27.5 – 32.8)	27.5	32.5	0.046
Hookworm	36.3 (32.8 – 39.9)	39.4	33.6	0.01
Bacterial infections				
Brucella spp.	0.6 (0.2 – 0.9)	-	-	-
Coxiella burnetii	2.2 (1.5 – 2.9)	2.5	1.9	0.32
Mycobacterium spp.	8.2 (6.8 – 9.6)	7.8	8.5	0.64
Viral infections				
HIV	5.3 (4.2 – 6.3)	2.9	7.3	<0.001

[1] Adjusted based on the complex survey design, see main text.
[2] On basis of Wald statistic. Not performed for the very rare events with few positive observations.

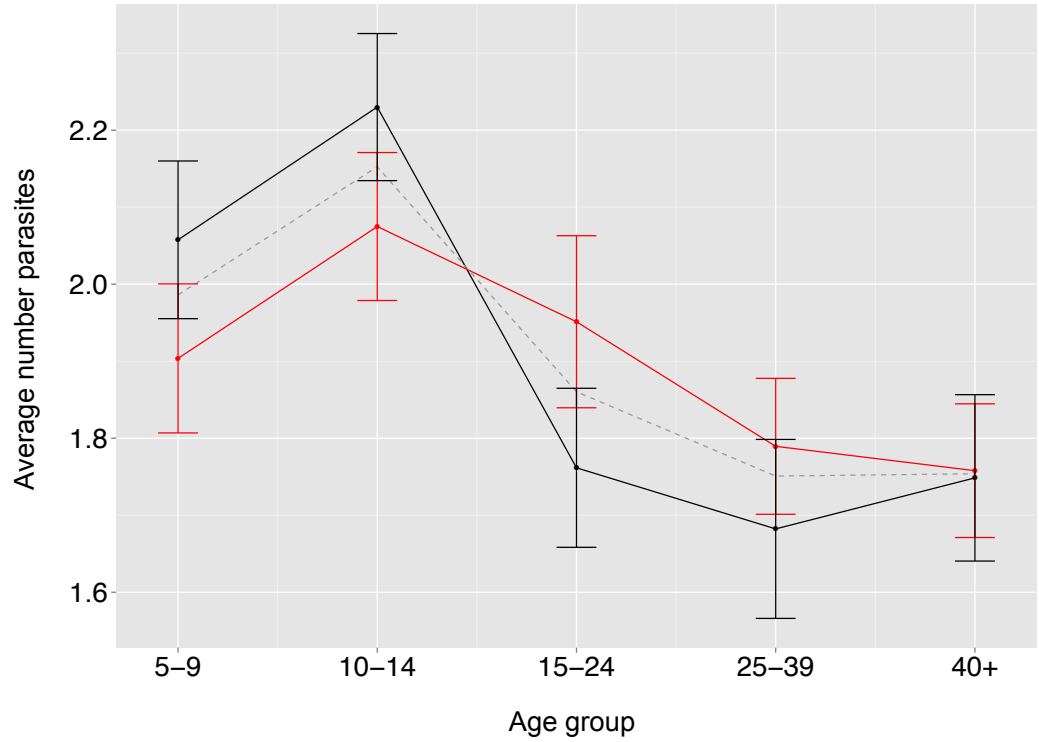


Figure 4.2. Age- and gender-stratified average parasite species count. Black line shows male average; red line shows female average, with 95% confidence intervals (not adjusted for survey design). Dotted grey lines represent age-stratified average.

4.3.2. Household-level infection

The observed and survey-adjusted household prevalence is shown in Table 4.2. More than 65% of households had at least one individual testing positive for each of *P. falciparum*, *E. histolytica/dispar* and hookworm.

Table 4.2 also summarises the output from the hypergeometric group-level prevalence procedure. The performance of the diagnostic tests used to define individual infection status are described in the Appendix (A.4.2). In general, diagnostic sensitivity for the tests used as part of this study was considered to be quite low, particularly for the diagnosis of gastrointestinal parasitism on the basis of a single faecal sample. By contrast, diagnostic specificity was generally considered to be reasonably high.

For outcomes that were rare and for which diagnostic sensitivity was particularly low, the observed household prevalence and that adjusted for diagnostic test error were substantially

different, with the latter being considerably higher in most cases. *Strongyloides stercoralis* provides a notable example: we expect that the diagnostic sensitivity for this parasite on the basis of a single faecal parasite to be less than 50%, and probably around 40%. In the case of household infection with HIV, for which diagnostic sensitivity and specificity is extremely high, the observed and test-adjusted estimates are almost the same.

On the basis of the observed data (i.e. unadjusted for test error), the average species richness (i.e. unique species) per household was 4.7 (95% CI 4.4 – 4.9), with a range of 0 to 13, and median of 4. Only 14 (3.7%) households had zero infections, whilst 117 (31%) had 6 or more, 77 (20.1%) had 7 or more, 37 (9.8%) had 8 or more, and 16 (4.2%) had 9 or more unique infectious agents in the people living within it. There was a strong relationship between the number of people sampled and the number of unique species detected within a household ($r = 0.61$, $p = <0.001$).

Table 4.2. Observed (unadjusted) estimates of household prevalence, those following adjustment for the complex survey design, and the output from the BDFree (group prevalence) procedure with adjustment for test error.

Infection	% (unadjusted) (95% CI)	% (survey adjusted) (95% CI)	% (BDFree) (Std. error)
Gastrointestinal parasites			
Balantidium coli	0.2 (0 - 1.6)	0.1 (0 - 0.3)	-
Fasciola spp.	0.2 (0 - 1.6)	0.2 (0 - 0.5)	-
Entamoeba hartmanni	0.2 (0 - 1.6)	0.2 (0 - 0.7)	-
Endolimax nana	0.4 (0 - 1.9)	0.4 (0 - 1.0)	-
Hymenolepis spp.	1.0 (0.3 - 2.6)	0.7 (0.05 - 1.5)	-
Taenia spp. (eggs)	1.0 (0.3 - 2.6)	1.2 (0 - 2.6)	-
Blastocystis hominis	1.7 (0.7 - 3.6)	2.4 (0.7 - 4.0)	-
Cryptosporidium spp.	2.7 (1.4 - 4.8)	2.9 (0.9 - 4.9)	27.0 (0.5)
Strongyloides stercoralis	13.4 (10.3 - 17.1)	12.7 (9.0 - 16.5)	61.0 (0.7)
Giardia spp.	12.8 (9.8 - 16.5)	14.6 (10.3 - 19.0)	33.7 (1.1)
Taenia spp. (HP10-ELISA)	17.8 (14.3 - 21.9)	17.3 (12.9 - 21.8)	-
Schistosoma mansoni	19.0 (15.4 - 23.2)	14.3 (10.6 - 18.0)	37.7 (1.4)
Trichuris trichiura	30.4 (26.0 - 35.2)	31.2 (26.5 - 35.8)	45.7 (1.6)
Ascaris lumbricoides	23.8 (19.9 - 28.3)	30.3 (24.7 - 35.9)	41.0 (1.5)
Iodamoeba butschlii	44.4 (39.6 - 49.4)	45.3 (39.6 - 51.0)	56.5 (1.7)
Taenia spp. (Copro-ELISA)	43.7 (38.8 - 48.6)	45.9 (40.0 - 51.8)	50.6 (1.6)
Plasmodium falciparum	67.1 (62.3 - 71.5)	68.2 (62.8 - 73.6)	70.1 (1.1)
Entamoeba histolytica	69.3 (64.6 - 73.7)	71.0 (66.0 - 76.1)	74.8 (1.7)
Hookworm	69.6 (64.8 - 73.9)	72.3 (67.2 - 77.4)	77.9 (1.5)
Bacterial infections			
Brucella spp.	2.9 (1.6 - 5.1)	2.6 (1.1 - 4.1)	7.8 (0.6)
Coxiella burnetii	11.1 (8.3 - 14.6)	10.2 (6.1 - 13.7)	22.0 (0.9)
Mycobacterium spp.	30.3 (25.8 - 35.2)	33.7 (28.1 - 39.3)	49.3 (1.5)
Viral infections			
HIV	21.9 (18.1 – 26.2)	20.8 (16.5 - 25.1)	21.4 (1.9)

4.3.3. Spatial distribution of household infection

Spatial heterogeneity in household risk of infection was observed for all of the pathogens under study, as shown in Figure 4.3 for HIV, *P. falciparum*, TB and *S. mansoni*, Figure 4.4 for the soil transmitted helminths and *Strongyloides*, Figure 4.5 for the gastro-intestinal protozoa and Figure 4.6 for the zoonotic infections. As might be expected given the difference in household prevalence when comparing the observed data with that derived from BDFree procedure (Table 4.2), the average smoothed risk estimates were generally

higher across the whole of the study area than those generated on the basis the single smoothed surface using observed data. This was particularly profound for the rarer outcomes, such as *S. stercoralis* (Figure 4.6M to P). However, in general, accounting for household misclassification did not appear to have major effect on the general distribution of areas of elevated household risk across the whole range of estimates for each infection (Fig 4.3, 4.4, 4.5 and 4.6).

The kernel smoothing procedure was more sensitive in identifying areas of significantly elevated relative risk than the Bernoulli spatial scan statistic. The former identified areas of elevated relative risk at the 5% alpha level for all infections under study (Figures 4.7, 4.8, 4.9, and 4.10) (albeit in very small areas in some cases, such as *S. stercoralis*, Figure 4.8D), whilst the latter identified significant clusters of elevated relative risk for seven out of the 15 mapped infections (*viz.* HIV (RR = 2.63, p = 0.0025); hookworm (RR=1.42, p=0.027); *S. mansoni* (RR=5.79, p=<0.001); *P. falciparum* (RR=1.55, p=0.0019); *I. butschlii* (RR=1.78, p=0.0039); *A. lumbricoides* (RR=2.31, p=0.030); and *T. trichiura* (RR=2.41, p=<0.001 and RR=2.26, p=0.0015)). These are shown in Figure 4.11. The geographic location of clusters identified via the spatial scan statistic were broadly equivalent to those identified via the KDE relative risk procedure, although the shapes of the clusters of the latter were quite heterogeneous.

The risk of infection for several infections was very localised, most strikingly for *S. mansoni* in the southern tip of the study area where it meets Lake Victoria (Figure 4.7.D), but also for HIV (Figure 4.7.B), *A lumbricoides* (Figure 4.8.A), and *T. trichiura* (Figure 4.8.C). Although the risk of *P. falciparum* and hookworm was generally high in all parts of the study area, there were also large areas with substantially (and significantly) elevated relative risk (Figure 4.7.A and 4.8.B). This was less obvious for the other very common parasite, *E. histolytica/dispar* (Figure 4.9.B).

Spatial overlap in risk

The extent to which areas of significantly elevated relative risk overlap are shown in more detail in Figure 4.12. Whilst areas of significantly elevated relative risk for each infection (on the basis of the KDE procedure) covered most parts of the study area, these occur less often in the northern half, with some (entirely subjective) indication of concentration of elevated relative risk for several infections (perhaps excluding the gastro-intestinal protozoa) in the southern and south-eastern parts of the study area.

Notable areas of substantial overlap include that between HIV and *S. mansoni* and *T. trichiura*, and that between hookworm and malaria (Figure 4.12).

There was evidence for a positive correlation in the spatial distribution of risk for several of the pairs of parasites under study, specifically *E. histolytica/dispar* and hookworm ($r = 0.63$, $p<0.01$), *E. histolytica* and *I. butschlii* ($r = 0.55$, $p<0.01$), hookworm and *P. falciparum* ($r = 0.47$, $p<0.05$), hookworm and *S. stercoralis* ($r = 0.61$, $p<0.01$) and *A. lumbricoides* and TB ($r = 0.68$, $p<0.05$) (Figure 4.13).

The output from the 'probability of at least *n*' procedure provides further support to the suggestion of a possible geographical gradient in household infection risk for multiple infections (Figure 4.14). Virtually the whole study area has a probability of >90% of observing at least 4 infections (out of a total of 15 mapped pathogens) in a randomly selected household (note, the average count of 4.7 per household described in section 4.3.2. is the unadjusted count). This begins to decline in the northern tip of the study area when considering the probability of at least 5 and 6 infections, but a clear band of very high probability (>80%) of observing at least 6 infections in a randomly selected household persists in the southern and eastern parts of the study area, which remains consistently high relative to the probability in surrounding areas up to a probability of at least 10 infections being present in a randomly selected household.

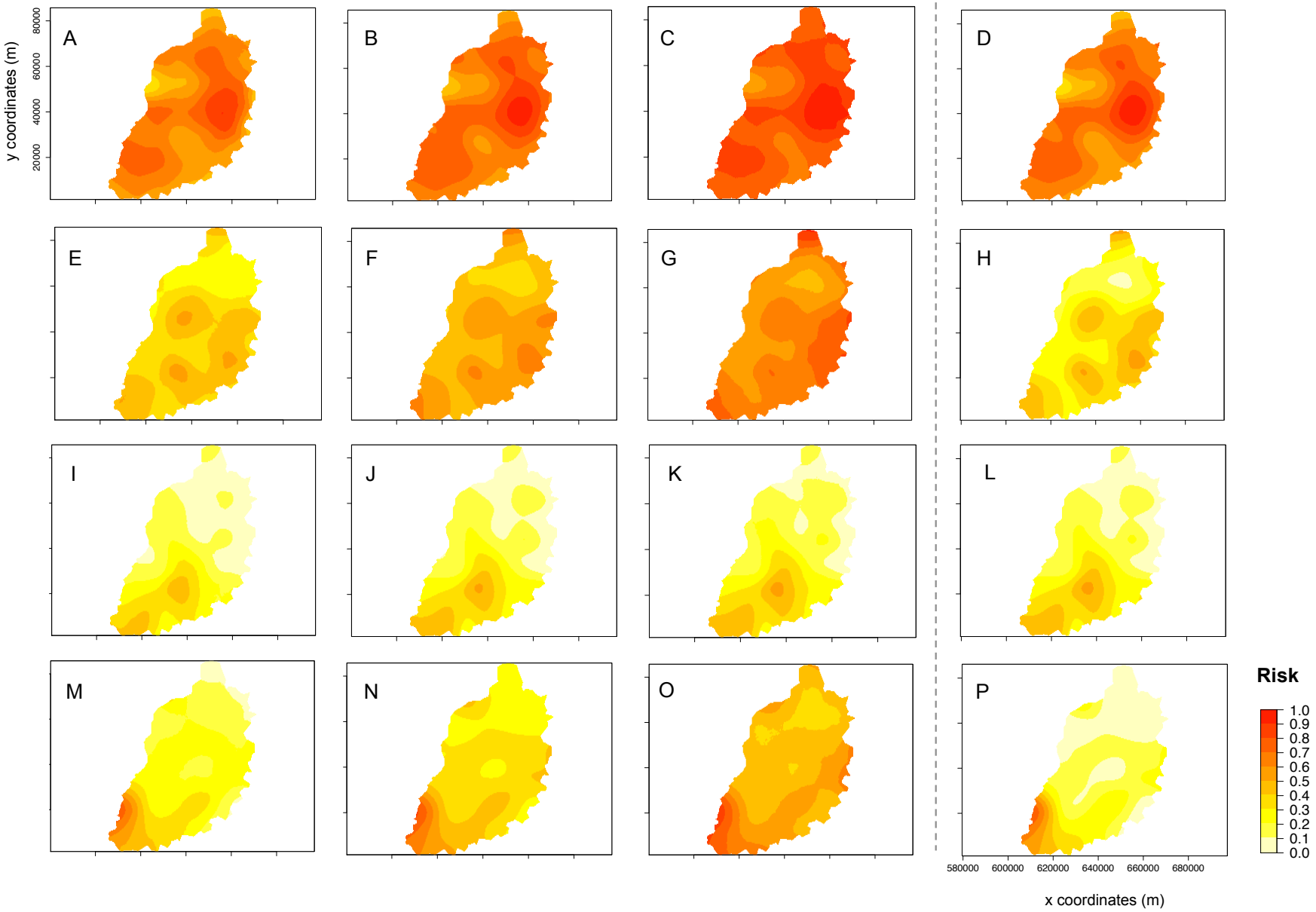


Figure 4.3. The 5th percentile (first column), average (second column), 95th percentile (third column) and observed (fourth column) household risk estimates for *Plasmodium falciparum* (A to D, respectively), TB (E to H), HIV (I to L) and *Schistosoma mansoni* (M to P)

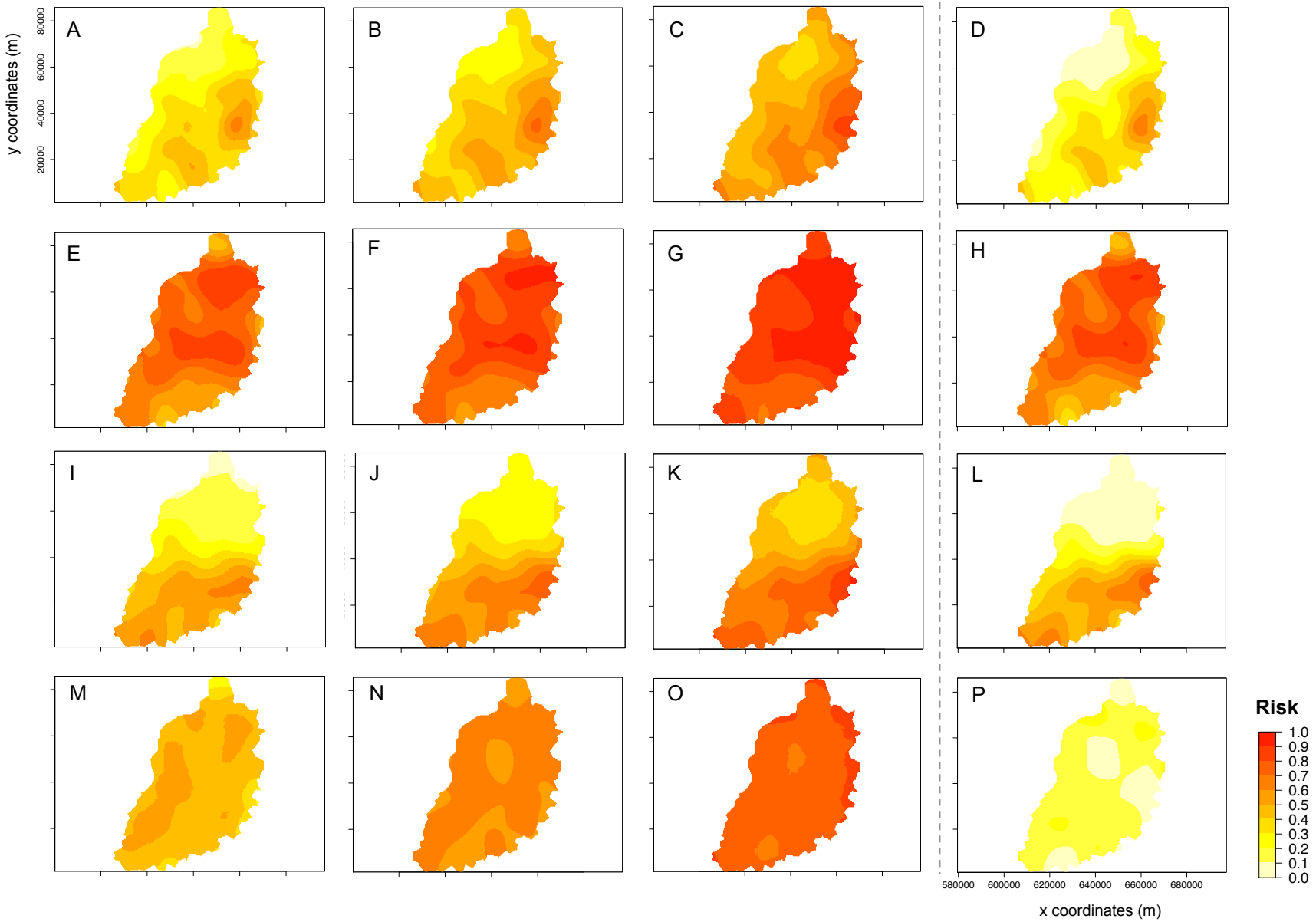


Figure 4.4. The 5th percentile (first column), average (second column), 95th percentile (third column) and observed (fourth column) household risk estimates for *Ascaris lumbricoides* (A to D, respectively), Hookworm (E to H), *Trichuris trichiura* (I to L) and *Strongyloides stercoralis* (M to P)

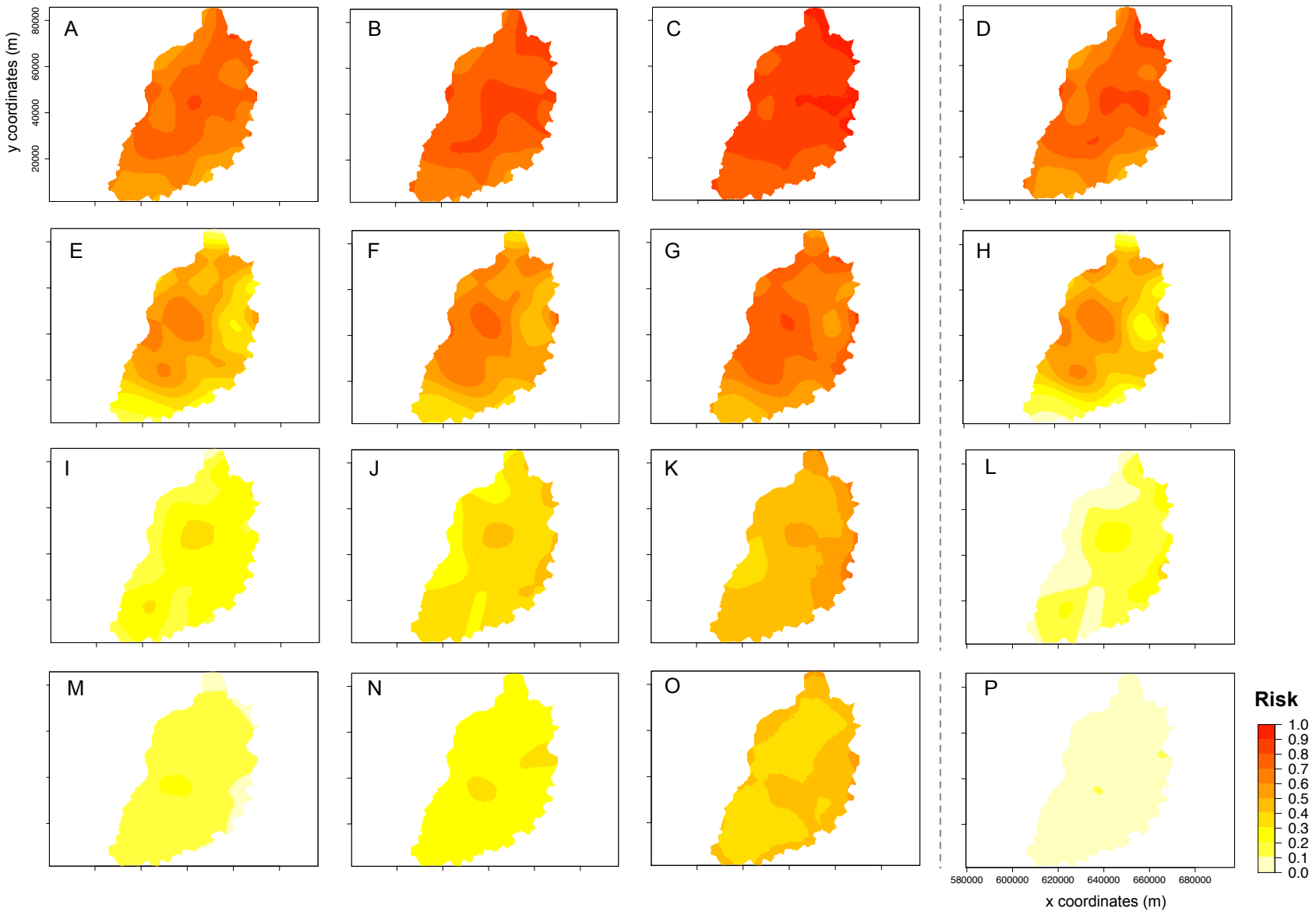


Figure 4.5. The 5th percentile (first column), average (second column), 95th percentile (third column) and observed (fourth column) household risk estimates for *Entamoeba histolytica* (A to D, respectively), *Iodamoeba butschlii* (E to H), *Giardia* spp. (I to L) and *Cryptosporidium* spp. (M to P)

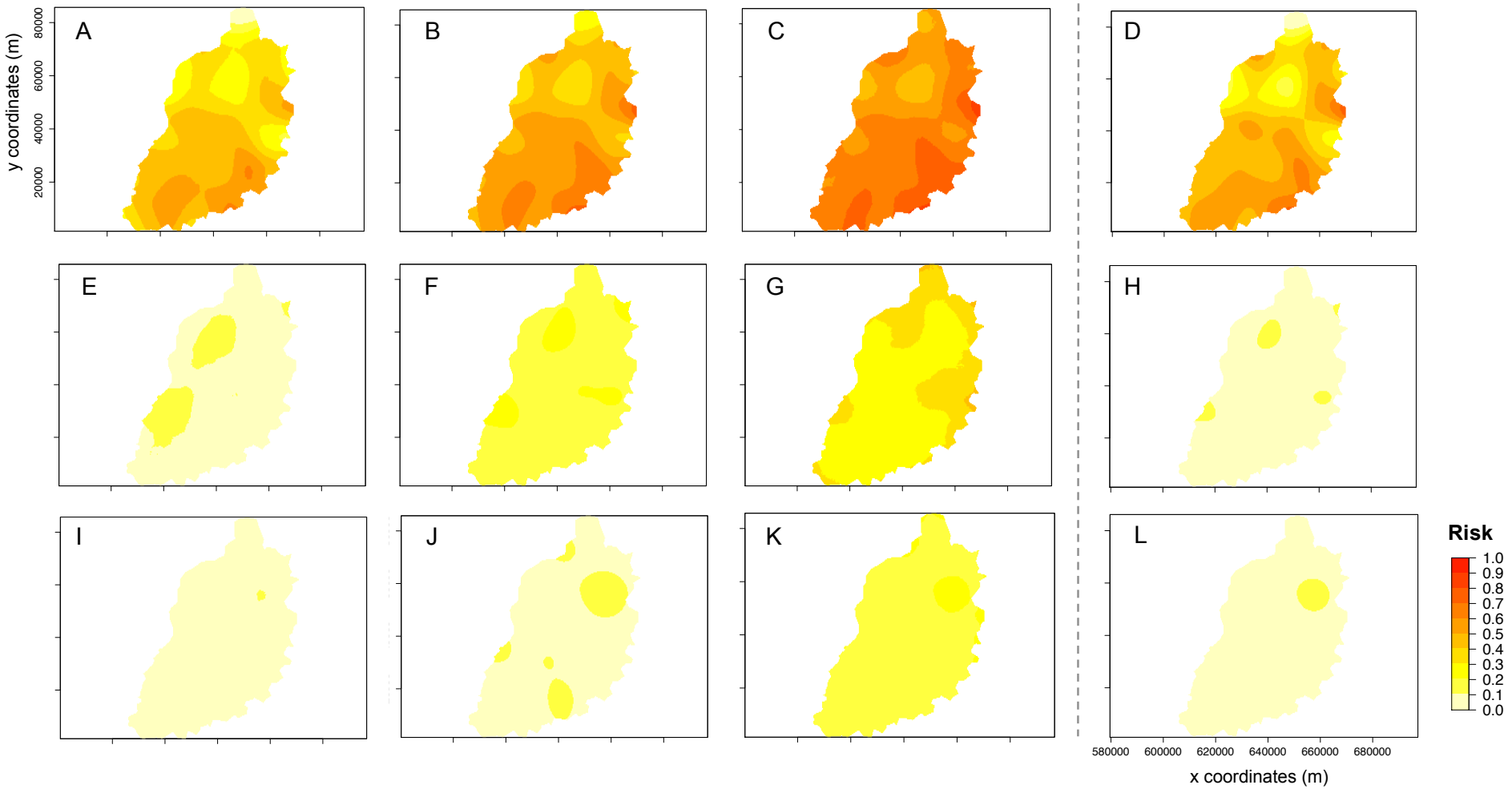


Figure 4.6. The 5th percentile, average, 95th percentile and unadjusted household risk estimates for *Taenia* spp. (A to D, respectively), Q-fever (E to H), *Brucella* spp. (I to L).

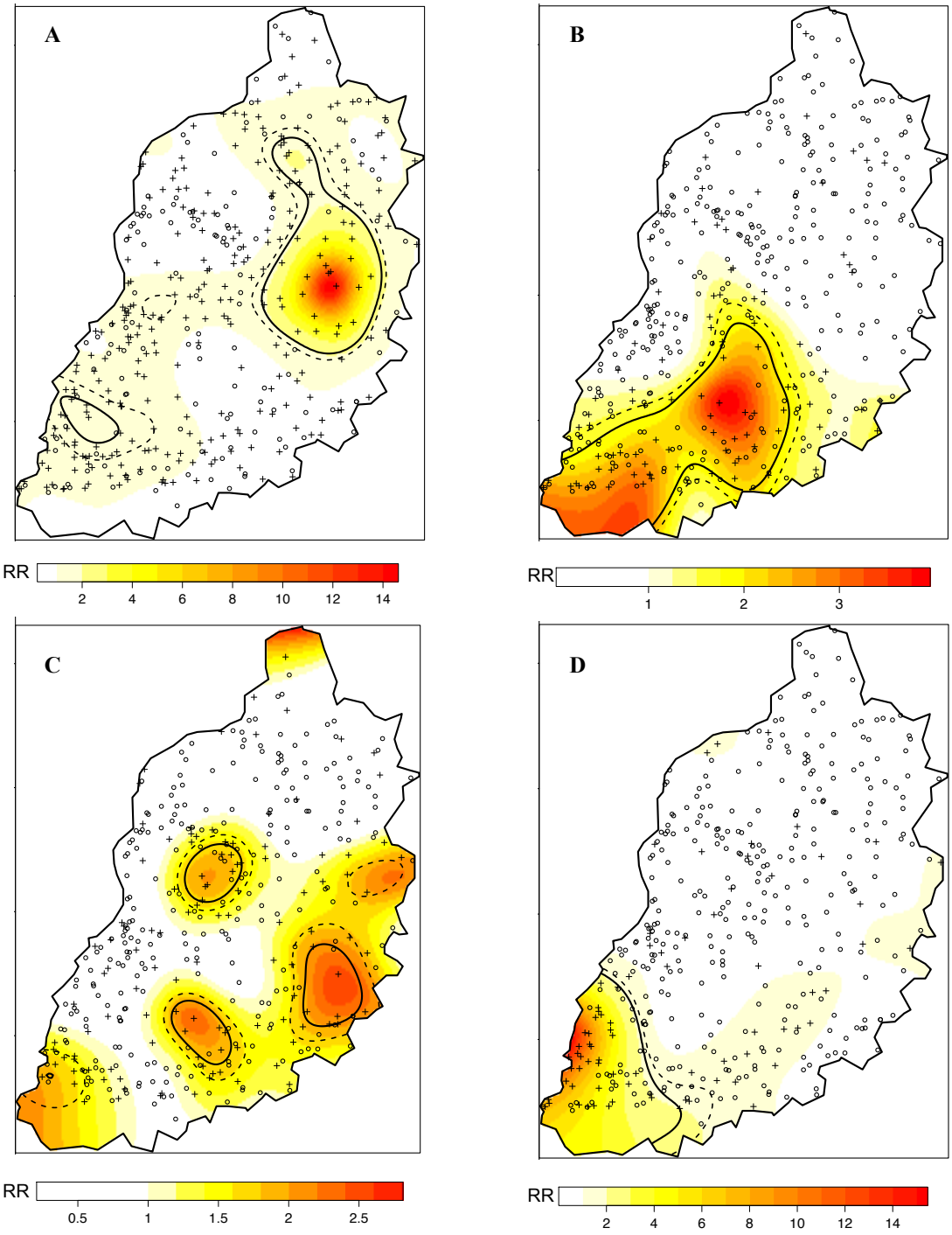


Figure 4.7. Household relative risk (RR) for *Plasmodium falciparum* (A), HIV (B), TB (C), and *Schistosoma mansoni* (D). Households with at least one infected individual are shown with (+), and those without as (o). Areas of significantly elevated risk at $p<0.01$ are shown by complete lines. Areas of significantly elevated risk at $p<0.05$ are shown by dotted lines.

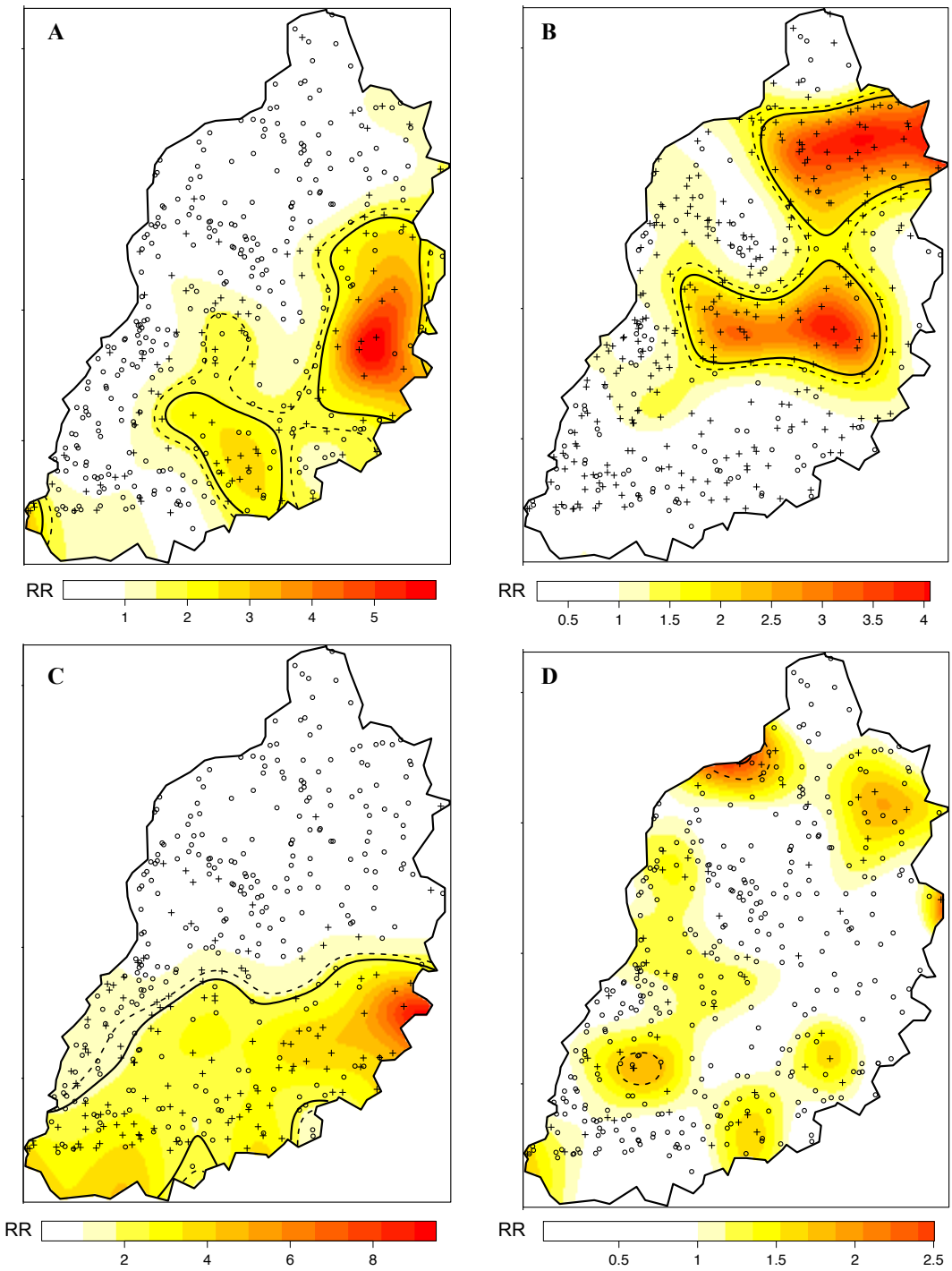


Figure 4.8. Household relative risk (RR) for *Ascaris lumbricoides* (A), hookworm (B), *Trichuris trichiura* (C), and *Strongyloides stercoralis* (D). Households with at least one infected individual are shown complete lines. Areas of significantly elevated risk at $p<0.05$ are shown by dotted lines.

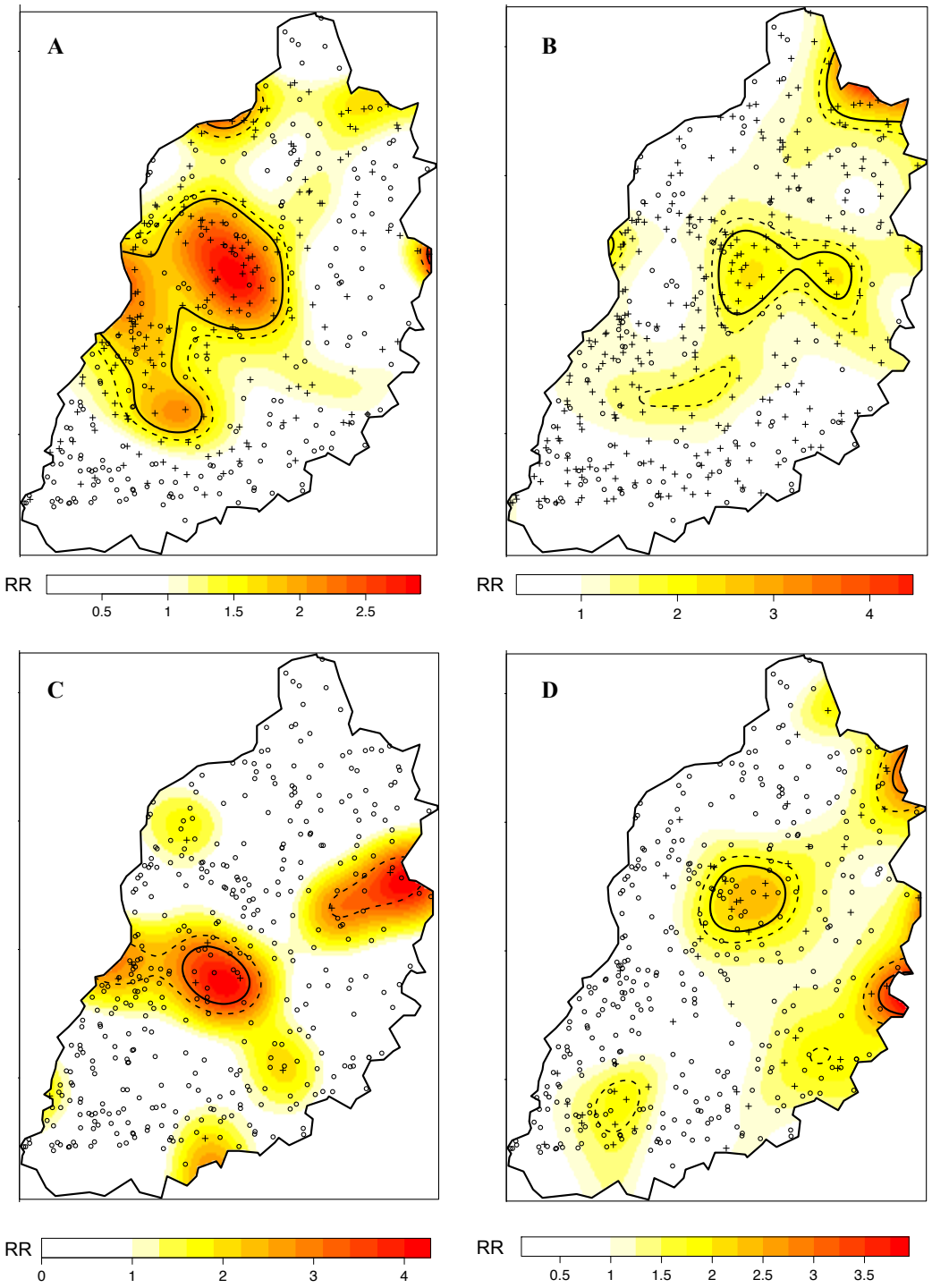


Figure 4.9. Household relative risk (RR) for *Iodamoeba butschlii* (A), *Entamoeba histolytica* (B), *Cryptosporidium* spp. (C), and *Giardia* spp. (D). Households with at least one infected individual are shown with (+), and those without as (o). Areas of significantly elevated risk at $p<0.01$ are shown by complete lines. Areas of significantly elevated risk at $p<0.05$ are shown by dotted lines.

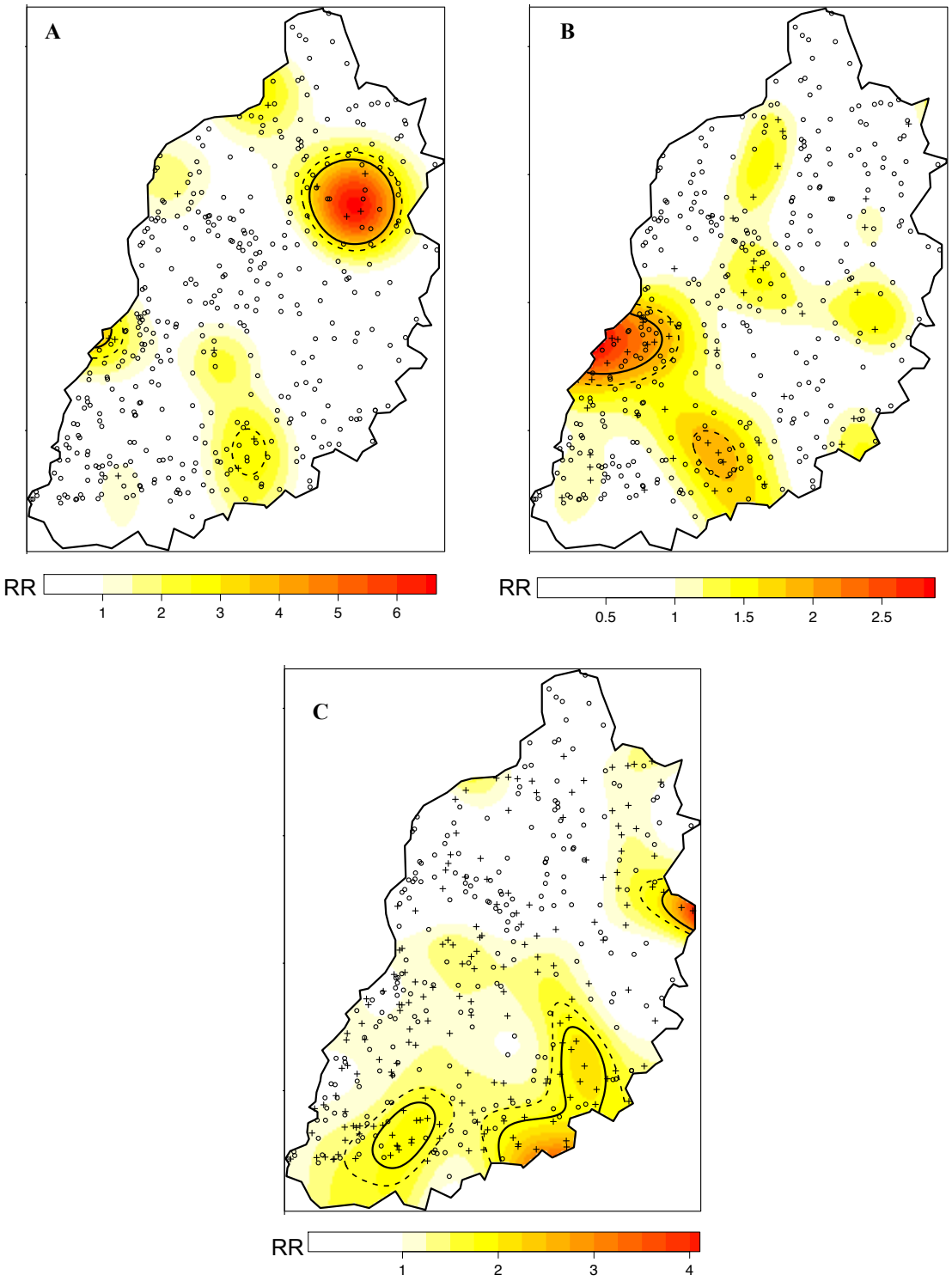


Figure 4.10. Household relative risk (RR) for *Brucella* spp. (A), Q-fever (B), and *Taenia* spp. (C). Households with at least one infected individual are shown with (+), and those without as (o). Areas of significantly elevated risk at $p<0.01$ are shown by complete lines. Areas of significantly elevated risk at $p<0.05$ are shown by dotted lines

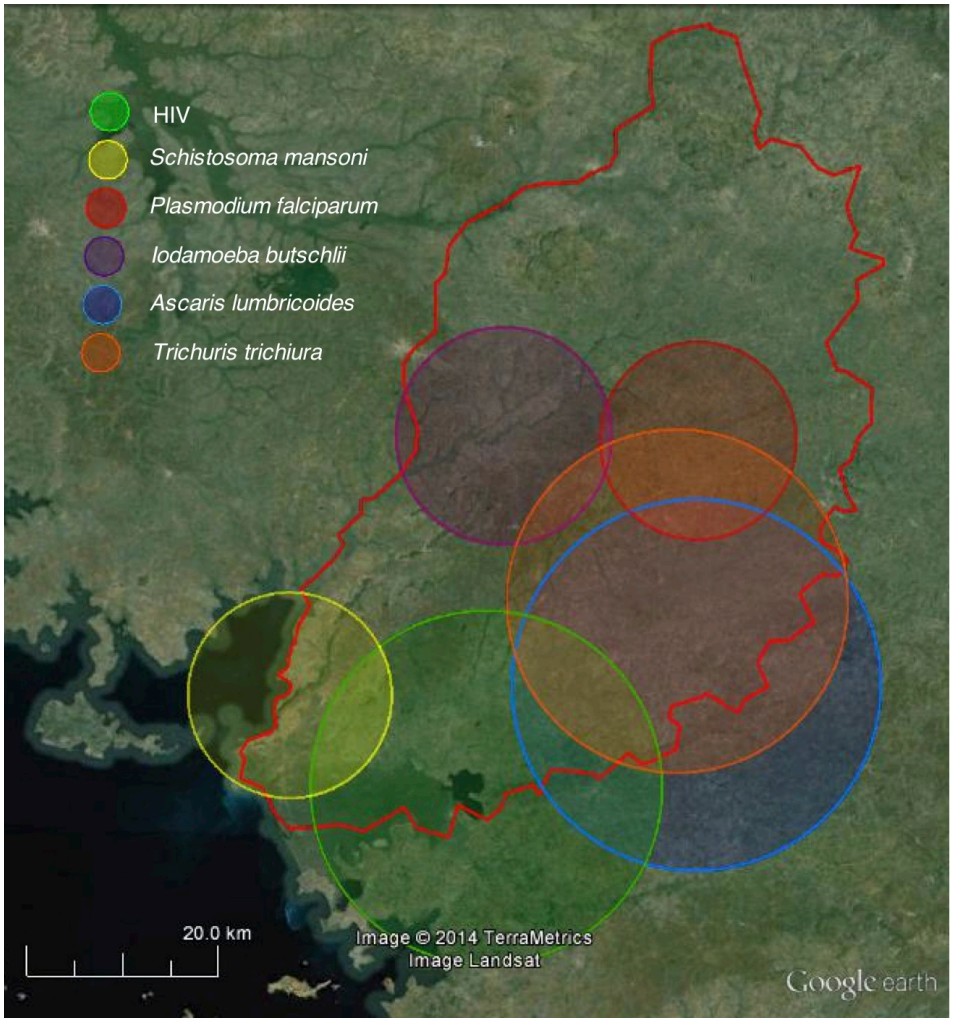


Figure 4.11. Statistically significant ($p<0.05$) clusters of elevated relative risk from the Bernoulli spatial scan statistic

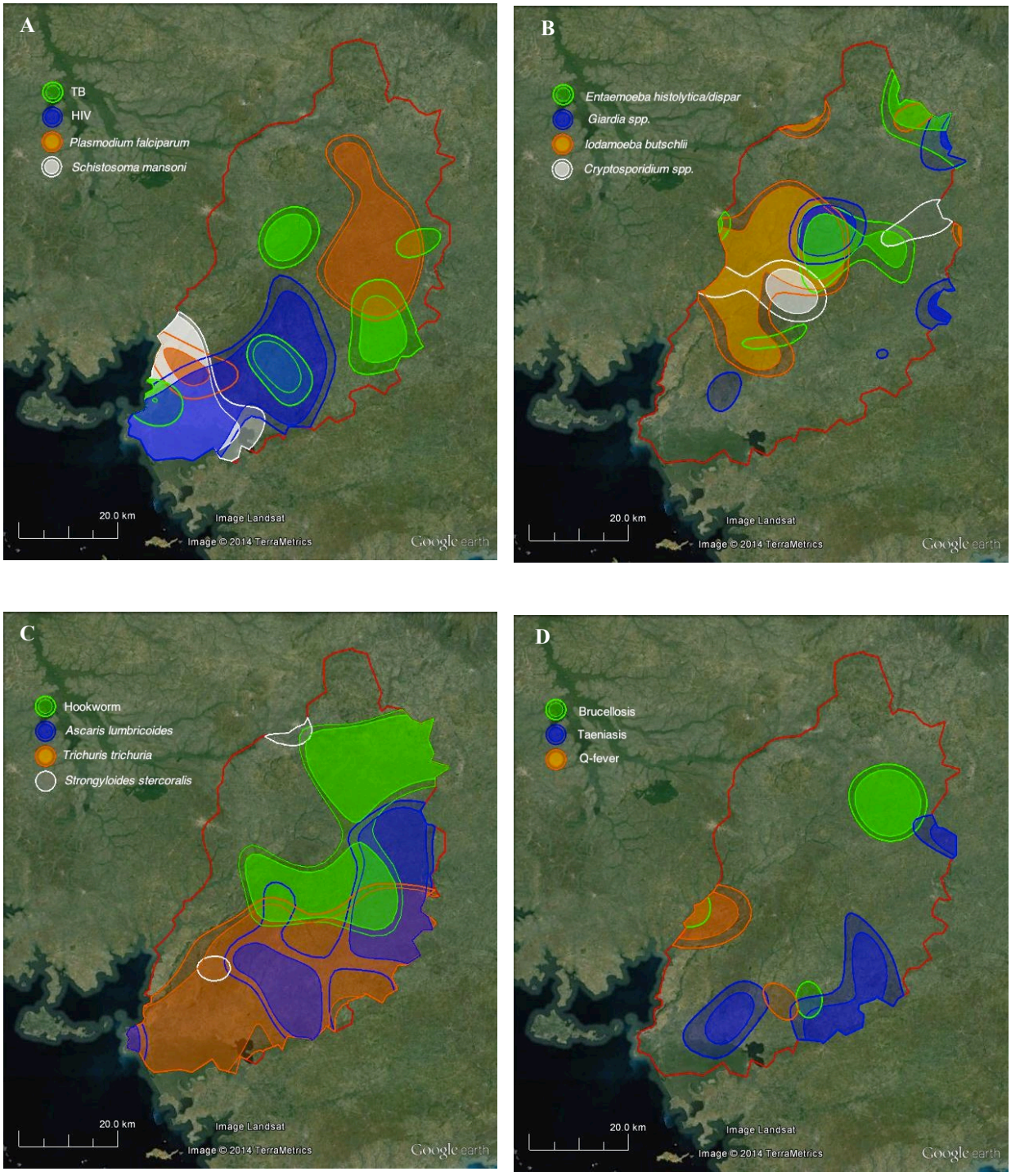


Figure 4.12. Relative risk contours for the different groups of infection (**A**: TB, HIV, *P. falciparum* and *S. mansoni*; **B**: *E. histolytica/dispar, Giardia spp., I. butschlii, Cryptosporidium spp.*; **C**: Hookworm, *A. lumbricoides*, *T. trichiura*, *S. stercoralis*; **D**: *Brucella* spp., *Taenia* spp., Q-fever). Inner (lighter) and outer (darker) contours represent relative risk that is significantly higher than the null at the 0.01 and 0.05 alpha levels, respectively. The dark region in the bottom left of each map is Lake Victoria.

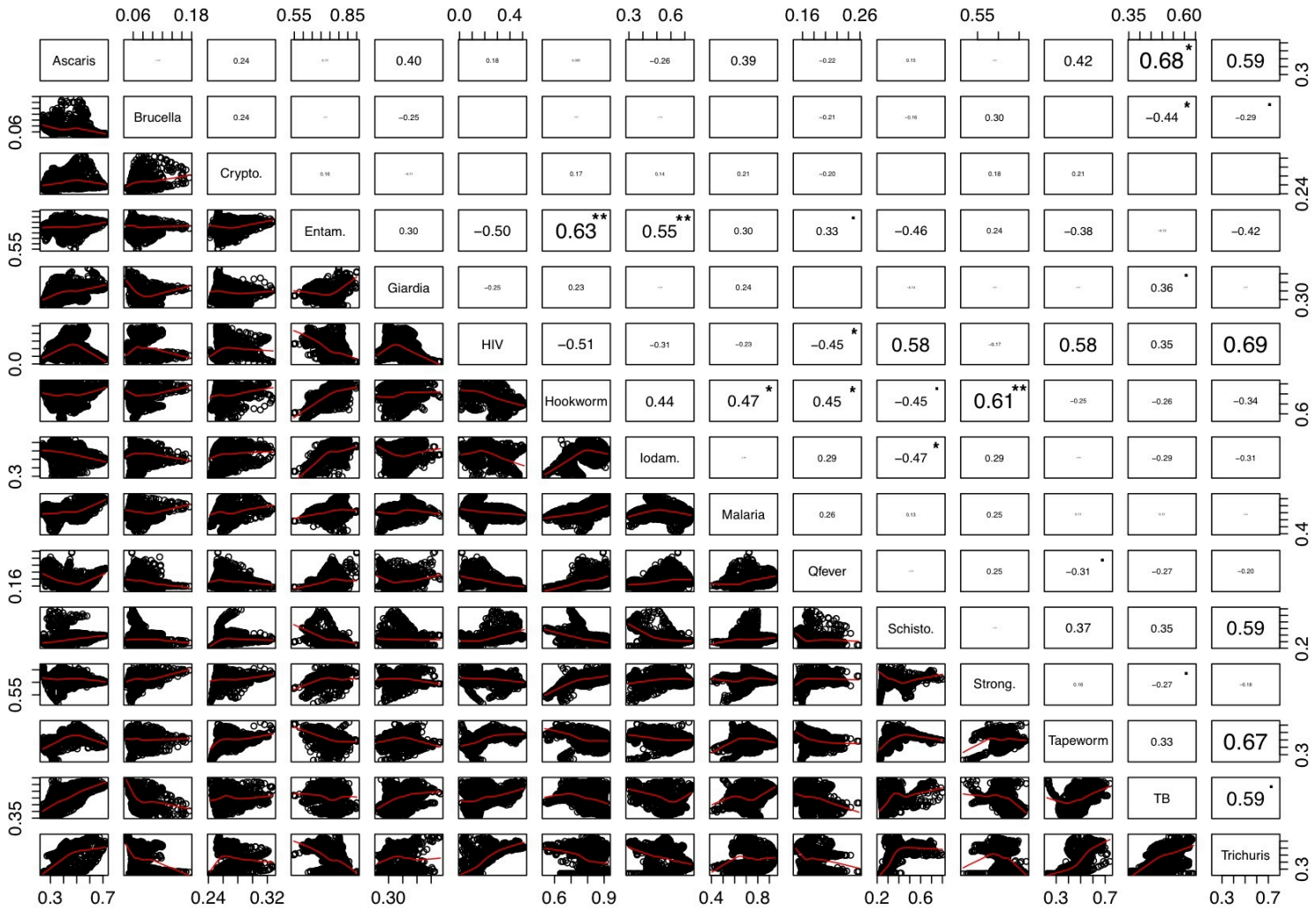


Figure 4.13. Correlation matrix comparing the average risk surface for each infection. The bottom left of the plot shows the graphical relationship between 1000 randomly selected points on the risk surface of each pair of infections shown in the names in the diagonal. The upper left of the plot shows the value of the correlation co-efficient, with size relative to the value. Statistical significance on the basis of sample sizes corrected for autocorrelation is given as **0.01, *0.05, °0.1

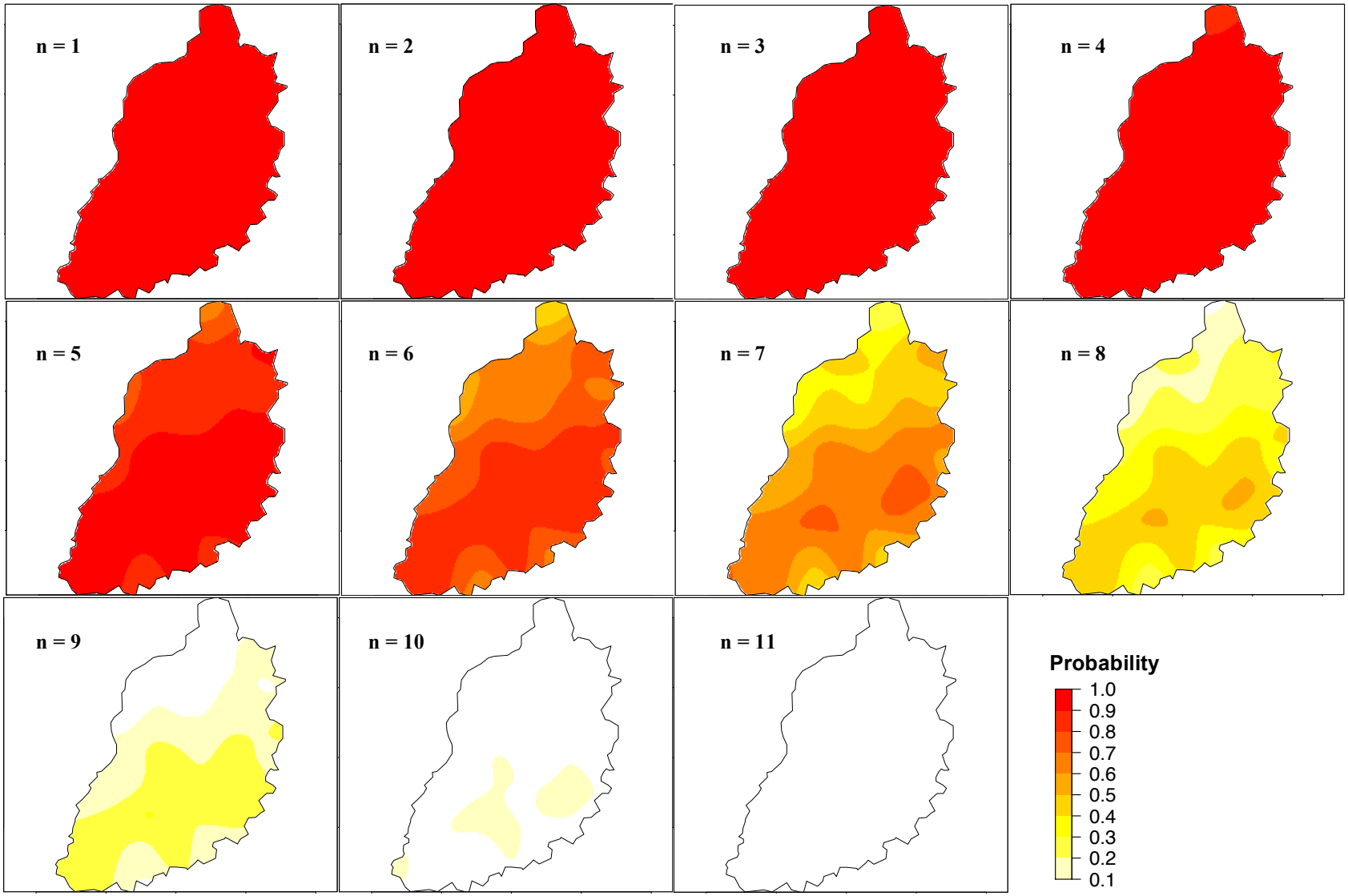


Figure 4.14. Probability of observing at least n out of 14 infections in a randomly selected household. There was negligible probability for n = 12 to 15, and these are not shown.

4.3.4. Spatial distribution of species richness

Households with the highest number of unique pathogen species relative to the number of people sampled were clustered within the southern part of the study area (Figure 4.15). This further supports the observation that there may be a geographical gradient in infectious disease risk (at least at the household level). The cluster of significantly elevated rates of infection considering only the gastro-intestinal parasites was broadly in the same location as the cluster of significantly elevated rates for all other parasites, and these overlapped to a large extent (Figure 4.15).

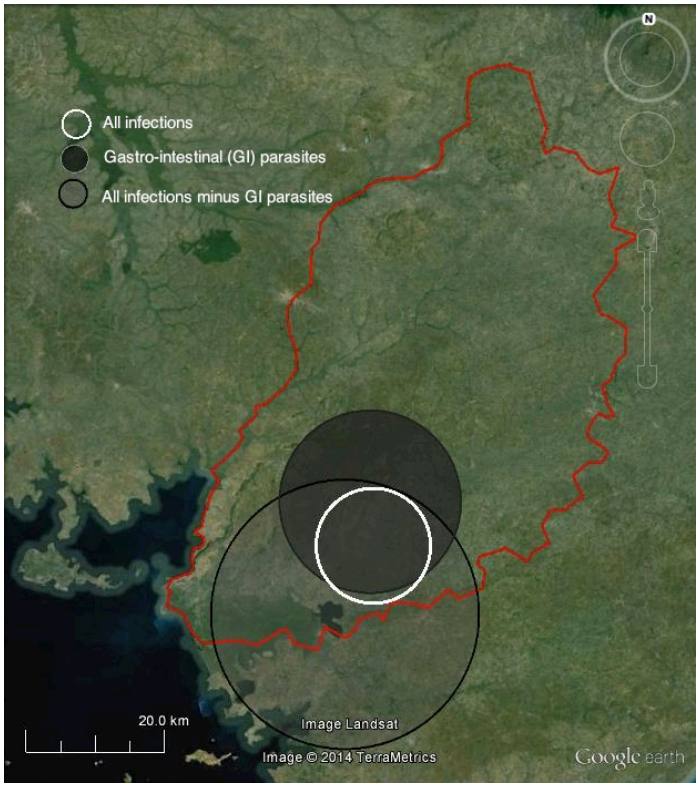


Figure 4.15. Statistically significant ($p<0.05$) clusters from the Poisson spatial scan statistic exploring infection counts.

4.3.5 Multilevel analysis

The output from the multi-level logistic regression analysis is summarised in Table 4.3. Between-division variation in risk of infection (of individuals in households) was observed for several of the infections. The variance was generally quite small, but in some cases, notably *T. trichiura*, and to a lesser extent HIV and *Taenia* spp. (defined using the results

from the copro-antigen ELISA only), the median odds ratios were greater than two. This suggests, for these infections, we would expect individuals in households in higher risk divisions to have (in median) two times or greater the odds of infection than those individuals living in households in lower risk divisions. Several of the infections had very low or zero variances at the division-level: a zero variance suggests the grouping of observations within divisions does not explain any of the overall variability in the infection risk, and households within divisions can be considered to be independent of a division level effect.

A degree of between-household variability was observed for all infections under study (those with an individual-level prevalence >5%) (Table 4.3). This was particularly profound for the soil-transmitted helminths, for which shared household-level effects (within the same division) explained 35% of the individual risk for hookworm, 54% for *T. trichiura* and 62% for *A. lumbricoides*. The median odds ratios were very high for these pathogens. To put the observed effects into context, in the case of *A. lumbricoides,* we might expect that for two people with the same individual characteristics randomly selected from different households in the same division, the person in a higher risk household would have (in median) nearly 9 times the odds of *Ascaris* infection. High levels of within household clustering were also observed for *Taenia* spp., and moderate levels for HIV.

The multi-level logistic regression model for *S. mansoni* was found to severely violate the assumption that model residuals are normally distributed (Snijders & Bosker 2011) and the results are therefore not presented. However, as can perhaps be expected given its very limited geographical distribution (Figure 4.5D), observations were highly correlated at the household level.

The output from the two-level Poisson regression model of household species counts showed minimal group-level variance at the division level (σ^2_D = 0.02, Median rate ratio = 1.2, ICC = 1.9%). Despite this, an interesting degree of spatial variability in division level residuals could still be observed. Whilst the effect sizes are small and confidence intervals generally wide, two divisions in the south and south-east of the study area had residual rates that were significantly greater than average, and there were two divisions in the northern half in which residual rates were significantly lower (Figure 4.16). This follows the general pattern of a geographic gradient in infectious disease risk described in section 4.3.3.

Table 4.3. Results from the multi-level logistic regression analysis

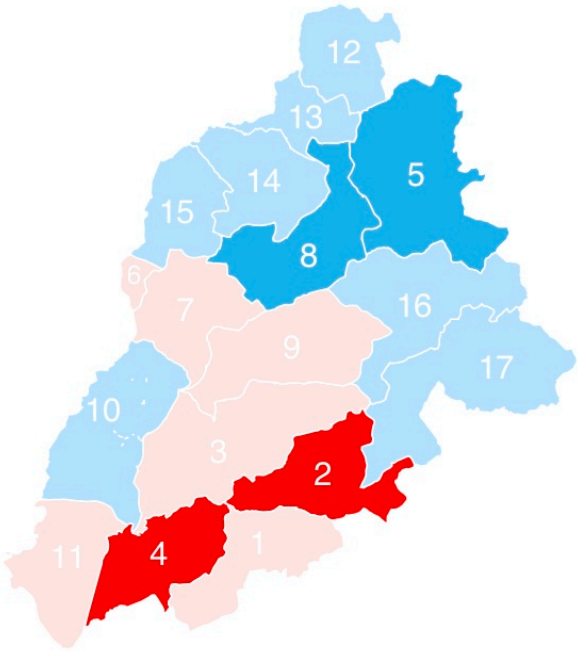
	Malaria	**Hookworm**	***Ascaris***	***Trichuris***
Fixed effects	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
5 – 9 years (ref)	-	-	-	-
10 – 14	0.84 (0.64-1.12)	2.02 (1.42-2.87)	0.72 (0.41-1.28)	1.79 (1.11 – 2.89)
15 – 24	0.23 (0.17-0.32)	2.01 (1.39-2.90)	0.51 (0.27-0.96)	1.31 (0.79 – 2.18)
24 – 39	0.13 (0.09-0.19)	3.04 (2.13-4.35)	0.26 (0.14-0.52)	0.61 (0.35 – 1.06)
40+	0.07 (0.05-0.11)	3.49 (2.47-4.93)	0.21 (0.10-0.43)	0.54 (0.32 – 0.86)
Male	1.07 (0.86-1.33)	1.41 (1.14-1.75)	0.80 (0.52-1.21)	0.61 (0.44 – 0.86)
Random effects				
σ^2_D (95% CI)[1]	0.03 (0-0.39)	0.24 (0.059-0.71)	0	2.39 (0.88 – 2.37)
σ^2_H (95% CI)[1]	0.51 (0.25-0.86)	1.56 (1.10-2.15)	5.3 (5.09 – 5.31)	1.48 (0.94 – 6.92)
MOR_D	1.19	1.59	1	4.35
MOR_H	2.01	3.57	8.93	6.48
VPC_D	0.84%	4.73%	0.00%	33.44%
VPC_H	14.2%	35.32%	61.76%	54.05%

[1] Estimated using profile likelihood methods (Hox 2010)

Table 4.3 (cont.).

	E. histolytica	***I. butschlii***	**Taenia *spp.***	**HIV**	**TB**
Fixed	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
5 – 9 years	-	-	-	-	-
10 – 14	1.36 (0.99-1.9)	1.33 (0.90-2.0)	1.13 (0.74-1.7)	2.62 (0.74- 9.2)	2.05 (0.78-5.4)
15 – 24	1.66 (1.20-2.3)	1.02 (0.68-1.6)	1.18 (0.75-1.8)	2.06 (0.55- 7.8)	5.09 (2.09-12.4)
24 – 39	1.12 (0.80-1.6)	0.84 (0.54-1.3)	1.09 (0.69-1.7)	17.66 (5.9-53.2)	9.88 (4.20-23.2)
40+	1.14 (0.83-1.6)	0.82 (0.55-1.2)	1.34 (0.88-2.0)	13.77 (4.6-41.5)	8.90 (3.82-20.7)
Male	0.75 (0.61-0.9)	0.82 (0.63-1.1)	1.31 (1.00-1.7)	0.50 (0.33-0.78)	1.32 (0.92-1.9)
Random					
σ^2_D	0.065	0.22	0.64	0.73	0.086
(95% CI)[1]	(0.004–0.26)		(0.23 – 1.73)	(0.08 – 1.91)	(0 – 0.48)
σ^2_H	0.60	0.61	2.37	0.50	0.41
(95% CI)[1]	(0.35 – 0.92)		(1.66 – 3.43)	(0.13 – 1.52)	(0 – 1.23)
MOR_D	1.27	1.56	2.15	1.96	1.32
MOR_H	2.16	2.38	5.24	2.86	1.95
VPC_D	1.6%	5.3%	10.2%	11.06%	2.28%
VPC_H	16.7%	20.1%	48.0%	27.17%	13.10%

[1] Estimated using profile likelihood methods (Hox 2010)



Rate ratio (95% CI)

1: 1.15 (0.96 – 1.37)
2: 1.25 (1.08 – 1.44)
3: 1.10 (0.92 – 1.30)
4: 1.18 (1.01 – 1.38)
5: 0.83 (0.71 – 0.98)
6: 1.00 (1.00 – 1.29)
7: 1.00 (0.84 – 1.19)
8: 0.87 (0.77 – 0.99)
9: 1.13 (0.98 – 1.31)
10: 0.99 (0.89 – 1.11)
11: 1.01 (0.87 – 1.18)
12: 0.95 (0.74 – 1.21)
13: 0.97 (0.79 – 1.18)
14: 0.91 (0.74 – 1.11)
15: 0.92 (0.78 – 1.08)
16: 0.99 (0.83 – 1.17)
17: 0.90 (0.75 – 1.08)

Figure 4.16. Distribution of division level residuals for the 2-level model describing count of infections at the household-level. Red divisions have a significantly elevated rates, light red divisions have a non-significantly elevated rates, blue divisions have a significantly reduced rates, light blue divisions have a non-significantly reduced rates.

4.4. Discussion

Infection with the aetiological agents of a range of infectious diseases is both common and widespread within this rural community in western Kenya, including infection with several of the 17 NTDs listed by the WHO (2012a) (*viz.* the STH, schistosomiasis and cysticercosis). Most people have multiple concurrent infections, as has been widely reported in other communities throughout the developing world (Buck et al. 1978). We clearly demonstrate that infection risk is not homogeneous across our study area, and that spatial heterogeneities in the probability of household infection exist to various degrees for each pathogen of interest. 'Hot spots' of significantly elevated relative risk were observed for almost half of the (selected) infectious agents under study on the basis of the spatial scan statistic, and for all 15 on the basis of the KDE (relative risk) procedure. This suggests some level of clustered transmission for most, if not all, of this wide range of pathogens within this moderately small area (3200 km^2).

The observed heterogeneity in risk raises potentially interesting questions about the factors that might influence transmission dynamics for each infection, and the obvious next step is to try to explain the effects observed. 'Landscape epidemiology' has traditionally focussed on environmental determinants of the spatial distribution of infection (Pavlovsky 1966; Lambin et al. 2010). For parasitic diseases, this has been fundamental to understanding the geographic limits of transmission (Pullan et al. 2014b), for predicting risk (Schur et al. 2011; Clements et al. 2010b; Pullan et al. 2011a; Clements et al. 2010a), and for understanding how climate change may impact upon changing distributions (Rogers & Randolph 2000). However, we can (generally) do little to change these broad environmental effects. Less often considered are potentially tractable contextual effects related to social inequality, such as community level socioeconomic status and educational attainment (Feldacker et al. 2011; Kayeyi et al. 2009; Chandola 2012), access to health care (Metcalf et al. 2014; Bates et al. 2004a), or infrastructural development, such as water and sanitation provision (Soares Magalhães et al 2011a; Pullan et al. 2014a). Such 'man made' determinants might be expected to impact upon a wide range of infectious diseases, including those with very different transmission routes. Hence, the suggestion of a broad geographic trend in risk for multiple infections within our study area, and particularly evidence for significant clustering of household pathogen species richness, may hint at the existence of important general contextual effects that certainly warrant further investigation. We explore some of these effects in the following chapters.

There is a very distinct hot spot of household HIV infection risk in the southern part of our study area. A recent call from UNAIDs was to 'know your epidemic, know your response' (UNAIDS 2014), which follows increasing recognition that there is not a single global HIV epidemic, but rather a multitude of diverse epidemics (Wilson & Halperin 2008). In this part of western Kenya, the HIV 'epidemic' appears to be extremely concentrated, such that the risk that a household contains at least one infected individual is 2 to 3 times greater within the identified hot spot (on the basis of the spatial scan statistic and the KDE procedures) than that outside it. Within the cluster identified by the spatial scan statistic, the individual-level HIV prevalence is 13.9% (95% CI 10.6 - 17.9) in all age groups, and 20.6% (95% CI 15.3 – 26.9) in people aged 15 to 64. This is considerably higher than the 2012 Kenyan national average in that age group (5.6% (KAIS 2013)), as well as that in the wider study area (9.3%), and clearly demonstrates the importance of considering spatial heterogeneity when considering HIV (and other) infection risk, even within small areas.

Ethnicity could be proposed as a possible explanation for the very high HIV rates observed in this region. The Luo community have been reported to have a particularly high burden of HIV, which is thought to be the result of traditional customs that use sexual practices as part of 'cleansing' rituals (Ayikukwei et al. 2008). The four main tribal groups in our study area (Luo, Luhya, Teso and Samia) are anecdotally quite spatially clustered, and the Luo community resides predominantly in the southern half of our study area, and particularly towards Lake Victoria, where HIV is also most prevalent. Indeed, and given the expected spatial dependency of the tribal groups, as well as the spatially structured nature of risk for most infections, it is likely that different tribal groups will be disproportionately affected by a range of different infections.

It is quite striking that the hot spot for HIV overlaps the cluster of elevated household polyparasitism (pathogen species richness) identified using the spatial scan statistic, as well as (much more broadly) the administrative divisions identified as having significantly higher residual rates in the multilevel Poisson procedure. Marked spatial heterogeneity in the distribution of HIV has been described by several authors (Cuadros et al. 2013; Zulu et al. 2014; Cuadros & Abu-Raddad 2014), but there has been very limited work to explore the co-distribution with other pathogens at a similar spatial scale to that explored here, and this has tended to focus on TB (Brunello et al. 2011; Musenge et al. 2013) or other blood-borne viruses (Zhou et al. 2014). The observed overlap of the area of high HIV relative risk with an area that seems to be at high risk for (household) polyparasitism could simply have occurred by chance (which we have not tested for), or the factors that impact upon elevated infection

risk for multiple pathogens in the identified "risky" area may also impact on elevated HIV risk. Alternatively, and as was described in chapter 2, parasitic (and other) infection may impact upon individual susceptibility to HIV, as well as on the infectiousness of HIV co-infected individuals. A possible hypothesis might therefore be that the high rates of some or all of the endemic infectious diseases observed within this 'high risk' area may have a positive impact upon HIV transmission dynamics. Much more work is needed to explore this potentially important effect, including control for the effect of ethnicity and other potential contextual and compositional confounders.

The spatial distribution of household risk of HIV correlated moderately strongly (but insignificantly) with that for *Taenia* spp., *S. mansoni,* and *T. trichiura*. As was described in the introduction, strong epidemiological relationships have also been demonstrated between HIV and *Schistosoma haematobium*, and although we didn't test for it as part of this study, this parasite is likely to be common in the swampy area in the southern part of our study area, and has been found to be less restricted to the shore of Lake Victoria than *S. mansoni* in other studies in the wider region (Sang et al. 2014; Lwambo et al. 1999). The co-occurrence of this parasite is therefore another potential explanation for the observed focality of HIV in our study area.

Additional spatial correlations of interest include the moderately strong (and significant) positive relationship between the smoothed surface describing household risk of TB and that for both *Ascaris* and *Trichuris*. As is described in chapter 3, immunomodulation associated with helminth infection has been hypothesized to increase susceptibility to TB, or progression to active disease. Regions with high helminth burdens also tend to have high TB burdens (Hotez et al. 2006a) but, to our knowledge, there has been no work to explore small scale spatial overlap in helminth-TB co-endemic communities.

We also observed a positive spatial correlation between smoothed risk surfaces for hookworm and *S. stercoralis*. Both species have similar transmission routes, that involve skin penetration, and might therefore be broadly influenced by the same environmental conditions (Becker et al. 2011). Males were at greater risk of infection for both parasites, which may also hint at shared exposure. The spatial distribution of household risk for hookworm was positively correlated with that for *E. histolytica/dispar*. We could find no previous reports of spatial overlap between these parasites, and Utzinger et al. (1999) found no relationships at the individual level in Cote d'Ivoire. However these (soil-transmitted and water-borne) parasites could reasonably be expected to co-occur in households and areas with inadequate sanitation. Several studies have reported positive associations between

individual risk of hookworm infection and malaria (Pullan et al. 2011b; Brooker et al 2006c; Brooker et al. 2012) and we found a moderately positive (and significant) association between these two parasites here.

In all cases, further exploration of the associations between these pathogens at both the individual- and household-level should be pursued. However, the results described here could be considered indicative of a potential relationship that might occur as a result of shared exposures or biological interactions. The use of smoothed risk surfaces has allowed us to demonstrate this spatial co-variation in a very straightforward way, and to use a simple hypothesis test that accounts for spatial autocorrelation. Other authors have derived, and compared, risk surfaces for multiple pathogens using more sophisticated geostatistical methods which provide better account for uncertainty as well as allowing the inclusion of covariates (Clements et al. 2010a; Soares Magalhães et al. 2011). Whilst building such models for the 15 pathogens described here would be a large undertaking, we could potentially explore doing so for those pathogens in which interesting relationships were observed using the KDE procedure. Multinomial models have also been used to explore co-occurrence of pathogen pairs within a spatial framework (Soares Magalhães et al. 2011b; Raso et al. 2006b; Schur et al. 2011), and the use of this approach could also be explored for co-infection at both the individual- and household-level.

Our findings have shown that place is important when considering the likelihood that an infection is present in a household. We have also shown that living in a high risk household has an influence on individual risk of infection, and that some degree of household clustering exists for all infections (although the 95% CI for household level variances are very close to 0 (the lowest possible bound) in the case of TB). This is hardly surprising given that we are dealing with communicable diseases, and those that can be transmitted within the shared domestic environment (Cairncross et al. 1996). Previous studies have reported substantial household clustering effect for the STH (Walker et al. 2011; Shapiro et al. 2005; Brooker et al. 2006a), and we repeat those findings here. Surprisingly few studies have quantified within household clustering for other infectious diseases within a sub-Saharan Africa context. We could identify no estimates of ICC/VPC for *Taenia* spp., but within household transmission is known to be important for this parasite (Lescano et al. 2009; Schantz et al. 1992). However, most of this work has focussed on within household transmission that results in cysticercosis (or the presence of encysted larvae) rather than taeniasis (or the presence of the adult worm in the small intestine). Here we show that taeniasis clusters at the household level (and did not test the effect for cysticercosis). This is perhaps not surprising

given that people in the same household will tend to share the same food (and taeniasis is a food-borne zoonoses: we describe the life cycle in more detail in chapter 7).

Quantifying ICCs can contribute to future surveys by informing design effects for sample size calculations (and therefore it is a shame they are not reported more often), but, when combined with household prevalence, can potentially provide useful information about the burden of an organism within a community. For those infections in which there was substantial within-household clustering (particularly the STH, but also taeniasis/cysticercosis and HIV), the proportion of "infected" households, or household prevalence, gives an indication of the proportion of people that are infected *and* those who are at high risk of infection. We did not account for diagnostic misclassification in our analysis of variance, but given the generally low sensitivity of the tests used it is quite likely that the degree of within-household clustering is even greater than that described here (Branscum et al. 2005).

By partitioning variance using these very simple models, we have been able to show that factors operating at the level of the household are important in explaining individual risk of infection. We attempt to identify what some of those factors might be in the following chapters.

The outputs from the 'group-prevalence' approach described here are best considered as an approximation of the true estimate of household prevalence (although of course that is unlikely to ever be perfectly known). In particular, and in the absence of account for the prior probability that a household has each infection of interest, household prevalence may be overestimated for the rare outcomes. This may occur due to the prior probability of household infection being low in these cases, however since this is unobserved, it is difficult to say to what extent overestimation may have occurred. Perhaps the best we can say, then, is that our estimates represent an estimated upper bound (which has its own uncertainty), and that we would expect the *true* household level prevalence to lie somewhere in between the observed and predicted. Since we believe household prevalence is a useful measure, the lack of a straightforward approach to derive accurate estimates that account for test misclassification when dealing with small group sizes suggests further methodological work is required. Zero-inflated models were introduced in chapter 2, and their use to derive household prevalence is further explored in chapter 6 and 7 of this thesis. However, these models require us to make additional assumptions about what "household infection" means in order to meet the requirements of binomial sampling, and may not be applicable to all infectious diseases (see chapter 6 for more discussion). Alternative solutions might be to use mixture models that model counts of infected individuals in "infected" households using a

Poisson (or similar) distribution. This would have the added advantage of allowing the incorporation of covariates within a regression framework. Stamey et al. (2008) described the use of Poisson models that allow the incorporation of diagnostic misclassification, and which have potential applications here.

Division was moderately important in explaining individual variation in infection risk for *Trichuris* and *Taenia spp.*, but in general, and despite evidence of some degree of spatial clustering for all infections (on the basis of the KDE procedure), the overall variance was quite low. This does not necessarily mean that each division has the same log odds for each infection, but rather that the clustering of households within division does not (in general) help explain a great deal of the overall variability in individual risk. Indeed, and as we observed for the residuals from the Poisson model, examining spatial variability in model residuals may still be useful in such cases. The lack of obvious variation in our relatively small number of divisions (17) may also be a consequence of the modifiable areal unit problem. This occurs due to the imposition of (uncorrelated) man-made boundaries on a process that is continuous on the landscape (Pullan et al. 2012; Pfeiffer et al. 2008). Given that spatial effects are clearly important for most of the common pathogens of interest, spatial analysis approaches that provide better account for the spatial dependency are certainly warranted (Chaix et al. 2005; Pullan et al. 2012; Magalhães et al. 2011).

The kernel smoothing approach described here provides a reasonably straightforward means with which to derive continuous risk surfaces. However, a considerable limitation is that is does not incorporate (or show) the associated uncertainty in those estimates. We attempted to display some of the error associated with household misclassification (as infected or uninfected) by using the probability of household infection from the hypergeometric herd prevalence procedure. The resulting distribution represents the (95%) range of possible estimates of risk across the surface of the study area in the presence of uncertainty about the household's true infection status. An additional, and extremely important, source of uncertainty that we have not accounted for in any way is that due to sampling error. Our moderately small number of households (416) can be considered to be quite sparsely distributed within the landscape of the study area (of 3200 km^2). Clearly, where sampling is sparse, confidence intervals for population averages can be expected to be quite wide.

Summary of main findings

1. We demonstrated that some degree of within-household clustering occurs for all pathogens under study, and that this was quite substantial in some cases. Considering household prevalence as a substantive outcome of interest may be useful for policy and planning;
2. We attempted to adjust household prevalence on the basis of diagnostic test misclassification, whilst describing the limitations and challenges of this approach;
3. We show that some degree of spatial heterogeneity exists in household-level risk of infection for all pathogens under study. Using a range of methods, we found evidence for a possible geographic gradient in infection risk for multiple infectious agents. This may indicate the existence of potentially important shared contextual and/or compositional effects for multiple infectious diseases within our study area.

Chapter 5

Exploring the social determinants of infectious disease risk in a poor farming community in western Kenya

5.1. Introduction

A social gradient in health has been shown to exist at virtually all of the societal and geographical levels in which it has been studied: typically, the higher the social position, the better the health (CSDH 2008; Marmot & Wilkinson 2005). These differences relate to the direct impact of material deprivation, as well as to the psychosocial effects of poverty (Marmot 2005). However, most of the evidence for this gradient comes from middle or high income countries. In low income countries, and in sub-Saharan Africa in particular, much less work has been done to explore the social determinants of health (but see Eshetu & Woldesenbet 2011; Fotso & Kuate-Defo 2005; Alvarez-Dardet 2000; Braveman & Tarimo 2002).

Understanding the social determinants of endemic infectious disease, which continues to be a major cause of morbidity and mortality throughout the developing world, is important not only for targeting policies that seek to make structural changes to reduce disease burden, but also for understanding the transmission dynamics of the infectious agents themselves. Whilst the neglected tropical diseases tend to afflict the poorest members of poorest communities (WHO 2012b), it may be the wealthiest individuals in the same communities who have the riskiest sexual networks (Kongnyuy et al. 2006; Awusabo-Asare & Annim 2008), who spend most time outside the home and therefore have greater exposure to a wider range of infections (Odone et al. 2013; Glynn et al. 2000), or whose food consumption practices put them at particular risk (e.g. regular consumption of meat contaminated by tapeworm cysts (Thomas et al *in prep*)). Understanding the social context of infectious disease risk, and particularly the potentially heterogeneous impact of socioeconomic position, becomes ever more important as integrated interventions are increasingly targeted towards multiple infectious outcomes (Kabatereine et al. 2010; Kolaczinski et al. 2007; Brady et al. 2006; Utzinger et al. 2009; Noblick et al. 2011).

In this chapter, we assess the extent to which a socioeconomic gradient can be considered to exist in a predominantly poor rural community in western Kenya, and quantify how individual infectious disease risk may be heterogeneously distributed across this gradient.

5.2. Methods and Results

Our analytical approach could be split into three sections:

1. Describe the household conditions in the study area;
2. Derive a locally appropriate index of household socioeconomic position (SEP);
3. Quantify the effect of household SEP on individual infectious disease risk.

We describe the methods and results for each section in turn.

5.2.1 Household conditions in the study area

To provide an indication of the average conditions in which people in the study area live, we summarised a selection of outcomes from the PAZ individual and household-level questionnaires (described in chapter 3). Where considered informative, we also performed simple statistical tests for association between variables using a univariable logistic regression with adjustment for sampling weights and household clustering (using the *svylogit* command in the *Survey* package in R (Lumley 2004)). The same test was used to assess the extent to which rurality, defined as Euclidean distance to an urban area, may also impact upon some of the household conditions of interest. 'Urban' was defined as an area with a population density greater than the 99th percentile for the study area on the basis of Worldpop (http://www.worldpop.org.uk/) population density data with a 100m resolution. Data were manipulated in ArcMap 10.1 (ESRI 2012, Redlands).

Almost all of the randomly selected households in this predominantly rural area could be considered to be agricultural, and 414 out of the 416 reported growing crops or owning at least one livestock species.

Buildings within household compounds were generally constructed from mud, dung and thatch (Figure 5.1), although most compounds (71.3%) included at least one dwelling with a roof made from metal or asbestos. A smaller proportion included a building with a cement or tile floor (23.9%) or walls made from cement or brick (17.9%). Building quality tended to be better in urban areas, and there was evidence for a negative relationship between the likelihood of a household containing a dwelling with an improved roof and distance to an

urban area (OR = 0.83 (0.72 – 0.96) per kilometre increase), but little evidence of an effect for improved floors (OR = 0.87 (0.73 – 1.03)) or improved walls (OR = 0.91 (0.75 – 1.09)).

The majority of households (76.6%) owned a latrine, although these varied in quality, with 4.7% being a simple open pit, 49.0% partially closed (a pit latrine walled by cut vegetation and without a roof), and 23% a fully enclosed and roofed structure. The majority of individuals reported 'always' using a latrine (67.4%), and only 4.4% reported 'never' doing so. Participants less than 19 years of age were less likely to report always using a latrine than adults (OR = 0.85 (0.82 – 0.89)). Even in households that owned any sort of latrine, 25.8% of residents reported not always (i.e. frequently, sometimes, never) using it. Households that were further from urban areas were less likely to own any latrine (OR = 0.78 (0.67 – 0.91)).

Households used a variety of water sources for their domestic needs, including well water (16.1%), boreholes (39.8%), springs (51.2%) and surface water (rivers, ponds, lakes: 22.7%). Relatively few had access to piped water (6.3%), and, unexpectedly, there was no obvious relationship between distance to an urban area and use of piped water (OR = 0.91, 0.70 – 1.16). Exploring this unexpected result revealed a cluster of households using piped water around Lake Victoria, a remote region in the south-west of the study area (see Appendix A5.1). There have been high levels of infrastructural investment in this region, including projects related to water supply (World Bank 2012). A small majority of households reported treating water (52.4%). Treatments included boiling, filtering or the addition of chlorine or iodine. Most households owned a bed net (91.8%), and there are on-going schemes in western Kenya to distribute these at high subsidy or free of charge (Githinji et al. 2010).

The use of mains electricity was rare (2.6%), and only a small proportion of households had access to any source of power (e.g. a rechargeable car battery, solar panels, a generator or mains electricity) (12.2%). There was weak evidence that households with access to any power source were more likely to be closer to an urban area (OR = 0.81 (0.62 – 1.02)).

Around half of all households reported having an external source of income in addition to home agriculture (51%). In most cases this was unskilled (e.g. fishing, sugar cane cutting, bicycle taxis), trade (e.g. small kiosks in local towns or villages) and more rarely in skilled work (e.g. teaching, nursing). An external source of income was more likely for those households that were closer to urban areas (OR = 0.85 (0.74 – 0.97)). Most adult members of households were minimally educated, with only 34.4% containing at least one adult with a secondary school education or above. Again, distance to an urban area was associated with adult education (OR = 0.82 (0.70 – 0.96)). There was a strong positive association between the education status of adults in a household and the likelihood of an external source of

income (OR = 4.8 (3.1 – 7.6)). Unexpectedly, 41% of households were headed by a female, although this was often (but to an unknown extent) due to a husband working away from home. Households headed by a woman were less likely to contain a secondary school educated adult (OR = 0.59 (0.38 – 0.89)), but there was no obvious difference in the likelihood of a source of external income (OR = 0.84 (0.57 – 1.24)), the presence of a latrine (OR = 0.83 (0.53 – 1.31)) or likelihood of treating water supplies (OR = 1.11 (0.75 – 1.65)).

Figure 5.1. A typical household in the study area. (*Photo: Charlie Pye-Smith/ILRI*)

*****

5.2.2. A locally appropriate index of SEP

Socioeconomic position (SEP)[3] has been defined as "the social and economic factors that influence the position individuals or groups hold within the structure of society" (Lynch & Kaplan 2000). It is the incorporation of these prestige-based, 'social' conditions, in addition

[3] SEP and socioeconomic status (SES) are often used interchangeably. We follow Lynch & Kaplan (2000) and use SEP throughout this thesis

to material goods, that makes it a difficult concept to define and to use empirically (Howe et al. 2012). In a rural sub-Saharan context, these factors may be related to the type and number of animals owned, the education status of family members, their employment, age, sex, religion, and ethnicity (Eshetu & Woldesenbet 2011). Together with a household's available material resources, these conditions might be expected to influence infectious disease risk directly by impacting upon exposure, or influence susceptibility to infection through a range of mediating factors. Understanding how such factors may be linked to health is important for deciding how SEP should be conceptualised (and incorporated into an epidemiological model) (Howe et al. 2008).

Following Boccia et al. (2011), we defined four 'domains' that we felt represented different elements of SEP in the community under study. The hypothesised contribution of each domain to SEP and to infectious disease risk (in general) is described below.

i. Household resources

An external source of income might be expected to impact upon a household's ability to purchase and invest in a range of health enhancing resources. This could be as straightforward as greater provision and consumption of nutritionally diverse foods (Kikafunda & Namusoke 2006), or contributions to improved sanitation (Jenkins & Scott 2007), greater access to medical services (Schellenberg et al. 2008) or to investment in health protection measures, such as bed nets and water treatment. Moreover, agricultural households with an external source of income may be more likely to make use of hired labour. This can provide household members, particularly mothers of young children, more time for health enhancing activities, such as food preparation and child care (Ukwuani & Suchindran 2003).

Hygiene practices and behaviours are likely to be related to education status (Basu & Stephenson 2005), and individuals who have attained higher levels of schooling can be expected to have better awareness of the safe handling of food and water, the disposal of waste and excreta, and of disease transmission pathways (Mata 1982).

Within the study area, and particularly amongst the Luyha community that make up the largest ethnic group, there is social prestige in large families (Kongstad & Mønsted 1980). In an agricultural society that relies on manual labour, a large number of adults, and in particular a larger number of adults relative to dependent children, may contribute to greater agricultural productivity and to reduced household economic vulnerability (Whitehead 2006). Conversely, smaller households may lead to less division of resources, and greater

investment in nutrition and education (Alsan et al. 2011). Female-headed households tend to have a high dependency ratio, and are often considered to be at the lowest end of a socioeconomic spectrum in rural communities (Charlton & Rose 2002). Indeed female-headed households in sub-Saharan Africa, even those that exist due to migration of a husband, tend to be smaller (i.e. have less acreage) and have less assets (Jayne et al. 2003). This, combined with the tendency for female household heads to have lower education than male household heads (IFAD 1999), may contribute to a lower health status amongst household members (Gakidou et al. 2010). In possible contradiction to this, the female control of household resources is often positively associated with improved child nutritional status (Ohiokpehai et al. 2007), and children from female-headed households are often less malnourished than those from male-headed households with access to equivalent resources (Kennedy & Haddad 1994). Nutritional status has been widely shown to influence susceptibility to infectious disease (Black et al. 2008; Saldiva et al. 2002; Hughes et al. 2004).

The duration that a household has been established could be expected to influence the position it holds within a community: longer established households may have more community support, and may be more likely to be surrounded by family members. Local support networks may contribute to reduced food insecurity (Charlton & Rose 2002) and to a greater ability to deal with or prevent ill-health (Zelner et al. 2012).

We collected data to describe each of these variables at the household level: an external source of income, regardless of source; whether any of the adults in the household had reached secondary school education; gender of the household head; total number of adults currently present; ratio of number of adults to number of children; and whether the household had been established for 5 years or more.

ii. Material resources

The ownership of durable assets and building quality is often considered to represent the consumption behaviour of a household (Filmer & Pritchett 2001). Consumption, rather than income (and income may a somewhat intangible concept in a subsistence economy) represents the long term economic position of an individual or household (Howe et al. 2012), which is likely to be important for most health outcomes, including infectious disease risk. In general, it could be expected that consumption behaviour will be linked to factors such as long term access to health care (which has point-of-use fees for most services in Kenya) as well as nutritional adequacy and diversity, both of which may impact on individual

susceptibility to infectious disease, and a household's ability to deal with it (Bates et al. 2004b).

Housing quality may be also impact directly on infection risk. In the sub-Saharan context, poor building quality may allow or increase the entry of disease vectors, including flies, mosquitoes, bats and rodents (Haines et al. 2013).

We asked household heads to report on their ownership of a range of durable assets (using closed questions), and interviewers observed and recorded the type of building materials used to construct all occupied buildings within the household.

iii. Household services

The likelihood of investment in household sanitary measures, such as latrine provision, is likely to be influenced by the education status of household occupants, as well the household's material resources (Jenkins & Curtis 2005) (although the monetary investment required to build a simple pit latrine need be minimal (Kar & Chambers 2008)). Households that invest in sanitary services are therefore likely to have both a higher SEP and reduced infectious disease risk. Latrine provision and access to a safe (i.e. uncontaminated) water source can be expected to directly reduce the risk of a range of infectious diseases and their sequelae (Ziegelbauer et al. 2012; Mwape et al. 2012; Esrey et al. 1991; Lin et al. 2013). Water from all sources can potentially be contaminated with pathogens in rural sub Saharan Africa (Hunter et al. 2009; Onda et al. 2012), but those sources that are subject to contamination from surface run-off (particularly rivers, lakes, and ponds, but also un-protected wells and springs) are likely to have the highest levels. Households with the highest SEP could be expected to invest in the best (least contaminated) water sources (such as piped water or boreholes) and to treat their supply.

Energy provision is widely linked with household welfare, and households that are 'energy poor' are often also income poor (Khandker et al. 2010). A source of power and the provision of light in the evenings may influence household productivity and child education, contributing to SEP, and therefore infectious disease risk indirectly.

Household heads were asked to report on water supply during both the dry and wet seasons, whether or not water is treated and by what means, and the type of power source available. The presence of a latrine, and its quality (in terms of construction), was recorded following observation.

iv. Productive assets

Animal ownership is often seen as a indicator of wealth in farming communities in sub-Saharan Africa (Randolph et al. 2007). Livestock can act directly as a financial reserve, with farmers seeking to sell some or all of their animal assets in order to meet unexpected financial demands (Kosgey et al. 2008; Dorward et al. 2005), as might occur due to ill health. Livestock may therefore act as an important buffer against adversity, and may give an indication of household vulnerability (Christiaensen & Subbarao 2005). We therefore considered the ownership of livestock assets represented a different element of household SEP to that indicated by the material resources (or consumption) domain.

Animal-source foods (ASF) are rich in a range of highly bioavailable micronutrients, including iron, Vitamin A and Vitamin B (Murphy & Allen 2003), and may be an important source of nutrition in livestock keeping households (Leroy & Frongillo 2007; Masset et al. 2012). Diets throughout sub-Saharan Africa, including those within the study area, are often dominated by carbohydrate-based starches, and supplementation with protein rich, nutrient dense ASF may have an important effect on health, and susceptibility to infectious disease (Hughes et al. 2004) as well as educational attainment (Neumann et al. 2007).

Despite these nutritional and financial benefits, livestock keepers and their families may also be at higher risk for a wide range of directly transmissible and food-borne zoonotic infections (Rwego et al. 2008; Randolph et al. 2007).

In the absence of sufficiently detailed data describing the home consumption of animal source foods, we focussed on livestock as assets. The local total monetary value of livestock assets owned by each household was summarised based on average estimates of value for each class and age of animal gathered from key informants at local markets and slaughterhouses (Okell 2011).

Total acreage of land owned by farming households could also be considered part of their 'productive assets', however these data were not collected.

5.2.2.1 Index construction

Whilst all of the individual variables described above could, in theory, be included in a regression analysis to explore their impact upon infectious disease risk, they are likely to be highly correlated which, together with the very large number of hypotheses to be tested, may lead to problems in estimation (Pickett & Pearl 2001). Alternatively, conditions such as education, a female household head or an external source of income could be chosen *a priori*

to be representative of broader SEP. However it is likely to be difficult to identify ideal measures that capture the multidimensional ways in which SEP may impact upon health (Vyas & Kumaranayake 2006; Howe et al. 2012). A useful alternative is therefore to employ a data reduction approach in order to derive a single measure or set of more parsimonious measures that describe the combined effect of these variables, and which can be used to represent latent SEP (Filmer & Pritchett 2001).

Such an approach has been used very widely to represent socioeconomic position in infectious disease epidemiology (Howe et al. 2012). In the majority of these applications, the index of SEP has been derived from the first principal component from a principal component analysis (PCA) of a set of variables describing asset ownership and housing quality (Filmer & Pritchett 2001) and, more rarely, characteristics of the household members (e.g. age, sex, occupation) (Cortinovis et al. 1993; Howe et al. 2012).

In the presence of the four distinct domains identified above, we felt a more natural method for multivariate data reduction was multiple factor analysis (MFA). This technique is an extension of (and analogous to) PCA, that is tailored to handle multiple groups of related variables. The approach involves first performing a separate PCA (or multiple correspondence analysis (MCA) in the case of categorical data) on each group of variables ('data tables'), with normalisation of the group by dividing all of its elements by the square root of the first eigenvalue from its PCA (or MCA when the data table is made up of categorical variables). The normalised data tables are then aggregated and analysed using a (non-normalised) PCA providing factor scores for the observations and loadings for the variables (Abdi et al. 2013). Thus, the approach takes account of the heterogeneity of biologically relevant groups of variables (Costard et al. 2009). Importantly (and in contrast to PCA), MFA assumes that there is an underlying causal structure, and that the co-variation is due to the presence of one or more latent variables (the factors) that exert a causal influence on the observed variables (O'Rourke et al. 2005). It is these latent variables that the outputs from factor analysis seeks to represent. Although multiple factor analysis has been used much less often than PCA to define SEP (Howe et al. 2012), it has tended to give similar results when comparing SEP indices derived by both methods (Sahn & Stifel 2002).

The MFA was performed in R using the package *FactoMineR* (Le et al. 2008). As three of the four groups of variables (the 'domains') described above (human resources; access to services; material wealth) were predominantly made up of categorical variables, we converted the only continuous variable (number of adults in the human resources domain) into quartiles in order to allow the initial steps of the MFA to be based on multiple

correspondence analysis (MCA) of grouped categorical variables in these three domains, rather than on PCA of a mixture of data types. Whilst PCA is the routine approach applied to derivation of a socioeconomic indices, the method was developed for continuous data and its use on categorical data (despite being very common (Filmer & Pritchett 2001)) potentially violates the distributional assumptions of the method (Kolenikov & Angeles 2004). Multiple correspondence analysis is the counterpart of PCA for categorical data (Greenacre & Blasius 2006).

All variables with a frequency of 5% or less were excluded from the MFA, since such variables may have an overly dominant effect on defining factors (Costard et al. 2009). In the presence of a reasonably large sample size (416 households), we performed no further variable reduction. One household had an unusually large total livestock value (£3969) and was excluded from the analysis as an outlier. The final set of categorical variables in the material wealth, access to household services and household resources domains, and their frequency in the study area, are shown in Table 5.1.

The fourth domain contained the total livestock value for each household. This ranged from 0 to £1593, with a median of £75 and mean of £155. Prior to the MFA, total livestock value (TLV) (the only continuous variable in the MFA) was scaled so that it had a mean of zero and standard deviation of one.

5.2.2.3 Results

The output from the MFA is shown in Figure 5.2. The x-axis of the plot (the first principal component (dimension 1) from the MFA) was considered (tentatively prior to validation) to be our index of SEP, with households with the highest value expected to have the highest SEP relative to all others in the study area. In the case of the MFA this dimension explained 13.3% of the total variation in component variables: the amount of variation explained by the first axis is always relatively low when dealing with binary categorical data, and a low value for the variance explained is not necessarily indicative of problem (Vyas & Kumaranayake 2006). The second principal component (dimension 2) from the MFA is included on the y-axis for graphical purposes only. The greyed points shown in the plot in Figure 5.2 represent the SEP score assigned to each household. The weights (scores, or principal co-ordinates) assigned to those variables that made a large contribution to the derivation of the index (i.e. contribution to inertia of an axis) are also presented. The position of these variables on the x-axis can be interpreted as describing their importance in defining the position of households along it (i.e. household SEP). It is therefore reassuring that this position matches the broad

idea of where they should be (i.e. the 'poorest' households do not own a mobile phone or have a latrine, whilst the 'richest' households have a power source and a fully enclosed latrine). The full set of scores assigned to each component variable used to derive the index, and their relative contribution, is given in Table A3.1 in the Appendix.

Table 5.1. The proportion of households owning each asset used to define the *material wealth* domain, with access to household resources described in the *access* to services domain, and the qualities in the household *resources* domain.

Material wealth	% (95% CI)[1]	**Access**	% (95% CI)[1]	**Resources**	% (95% CI)[1]
Motorbike	5.1 (2.7 - 7.5)	Piped water supply	6.3 (3.7 - 8.8)	More than 5 adults	16.8 (12.4 - 21.2)
Sewing machine	11.1 (7.6 - 14.5)	Source of electricity	12.2 (8.6 - 15.8)	3 adults	18.6 (14 - 23.3)
Improved walls	17.9 (13.4 - 22.3)	Well water supply	16.1 (11.9 - 20.3)	1 to 2 adults	32.0 (26.8 - 37.3)
Television	19.7 (14.8 - 25.5)	Closed latrine	23.3 (18.2 - 28.4)	4 – 5 adults	32.5 (27 - 38.1)
Improved floor	23.9 (19.1 - 28.6)	No latrine	26.6 (21.9 - 31.3)	Adults with sec. education	34.4 (28.6 - 40.1)
Watch	24.8 (19.4 - 30.2)	Borehole water supply	39.8 (34.6 - 45.1)	Source of external income	51.0 (45.1 - 56.9)
Phone charger	30.2 (24.7 - 35.7)	Partially closed latrine	50.2 (44.5 - 55.8)	Less children than adults	58.1 (52.5 - 63.7)
Clock	38.8 (33.2 - 44.4)	Spring water supply	51.2 (45.6 - 56.8)	Male household head	59.0 (53.2 - 64.8)
Cupboard	53.3 (47.4 - 59.2)	Water treatment	52.4 (46.9 - 57.9)	Established > 5 years	88.1 (84.2 - 91.9)
Torch	58.7 (52.8 - 64.7)	Bed net	91.8 (88.5 - 95.2)		
Sofa with cushions	61.8 (56.4 - 67.1)				
Bicycle	70.8 (65.3 - 76.4)				
Improved roof	71.3 (66.3 - 76.3)				
Phone	77.2 (72.2 - 82.2)				
Radio	85.2 (81 - 89.3)				
Bed frame	93 (90.2 - 95.8)				

[1] Point estimates and standard errors adjusted on the basis of the complex survey design (using the methods described in Chapter 4).

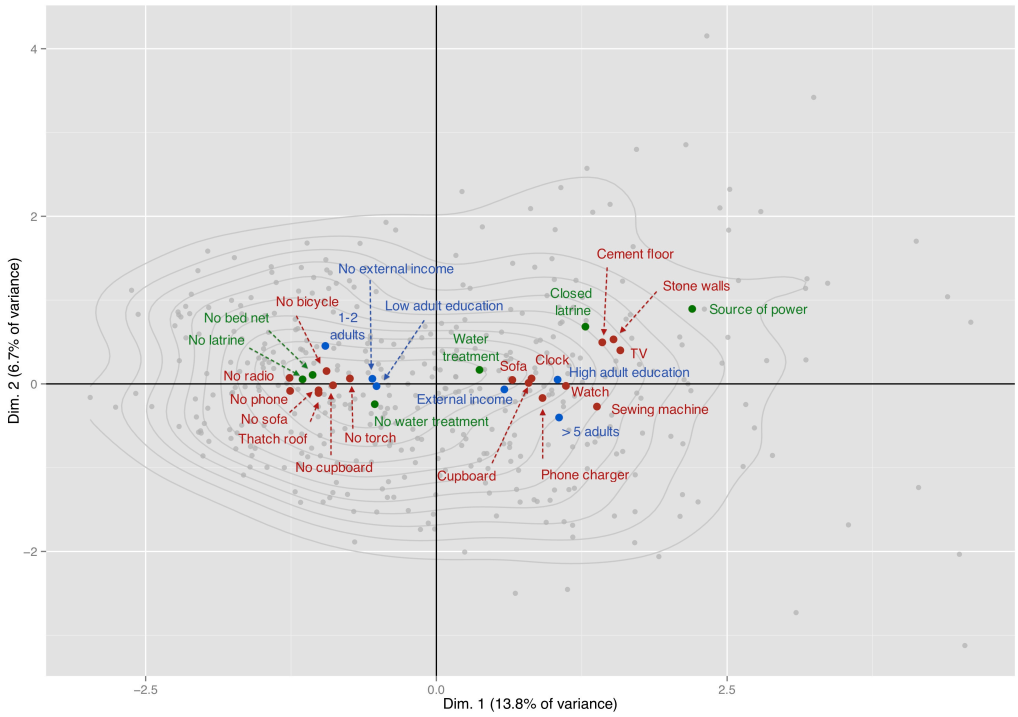


Figure 5.2. Scores of those variables which make a contribution to the inertia of first principal component (dimension 1) greater than 1, derived from the multiple factor analysis and used to represent **household socioeconomic status**. The greyed dots (and their density) represent the scores assigned to homesteads on dimension 1 (the index of SEP) and dimension 2 (included for graphical purposes only). Red variables are in the material wealth domain; green in the access to services domain; blue in the household resources domain.

A useful additional feature of the analysis of linked domains by MFA is that the results of the multiple correspondence analysis (MCA) of each of the 3 categorical domains (describing material wealth, access to services and household resources, and which forms the first part of the MFA procedure) can be extracted and examined. The first component from the MCA of variables making up the material wealth domain is shown in Figure 5.3, access to services in Figure 5.4 and household resources in Figure 5.5. The interpretation of these figures and the axes of the plots is the same as that for Figure 5.2, described above (i.e. households with the highest value on the x-axis can be considered to have the highest material wealth, access to services and household resources, respectively).

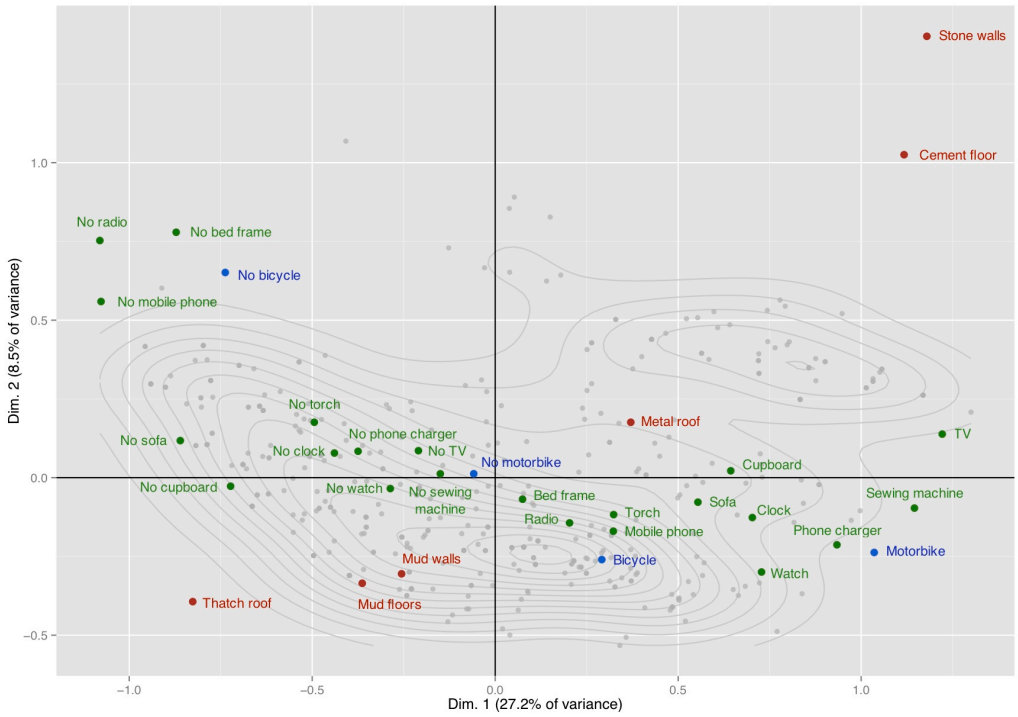


Figure 5.3. Scores of all variables used in the MCA to derive a proxy of **material wealth**. The greyed dots (and their density) represent the scores assigned to homesteads on dimension 1 and 2. Red variables are related to housing quality, blue to transport and green to household assets.

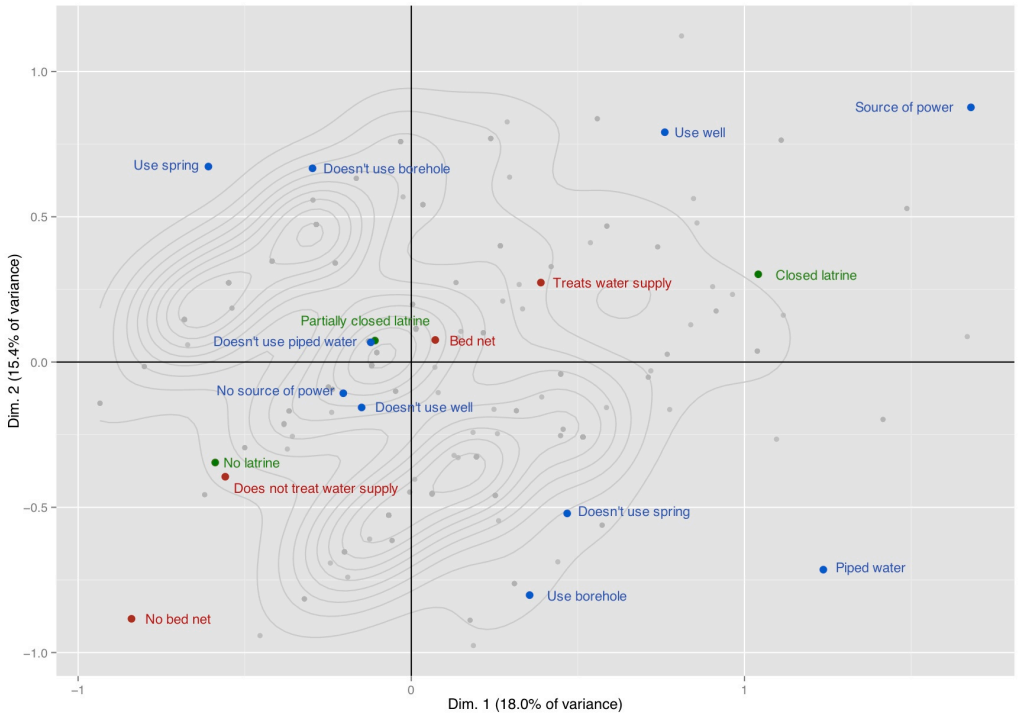


Figure 5.4. Scores of all variables used in the MCA to derive a proxy of **access to services**. The greyed dots (and their density) represent the scores assigned to homesteads on dimension 1 and 2. Red variables relate to health related variables, blue relate to water supply and power source and green to access to sanitation.

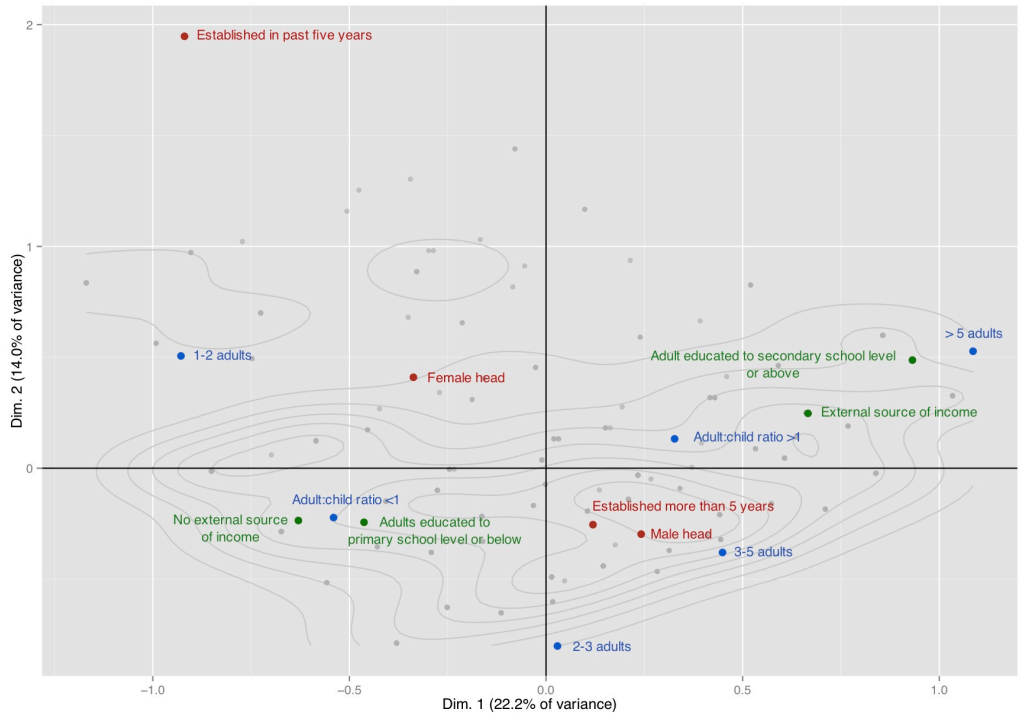


Figure 5.5. Scores of all variables used in the MCA to derive a proxy of **household resources**. The greyed dots (and their density) represent the scores assigned to homesteads on dimension 1 and 2. Blue variables relate to household size, green to education and income, and red to gender of head and duration established.

Validation

There is an element of the 'black box' to data reduction based on multivariate techniques such as MCA or PCA, and potentially even more so for MFA, where the complexity of the analysis is further increased. In order to ensure a degree of biological realism existed in the index of SEP derived from MFA, we compared the SEP score for each household with that derived using a subjective approach in which weights were assigned to variables on the basis of the analytic hierarchy procedure (AHP) described by Saaty (1980). Our use of this approach, which involved the pairwise comparison of each of the categorical variables used in the MFA (Table 5.1), is described in detail in Appendix A5.3.

The indices derived using the objective MFA and the subjective AHP were highly consistent (Spearman's rho = 0.94, $p<0.001$), suggesting a good degree of biological realism (or at least our perception of it) of the derived index of SEP used in the remainder of this chapter and throughout this thesis.

SEP in space

In Chapter 2 we observed substantial spatial heterogeneity in household-level risk of infection, as well as overlap in the spatial distribution of several infections. Since poverty is often considered to be at the root of much infectious disease risk (Farmer 1996; Alsan et al. 2011), it might be hypothesised that SEP is also spatially distributed in our study area, and that this spatial distribution might explain some of the geographic heterogeneity in disease risk.

To test the hypothesis that SEP is spatially structured, we examined the score for each household derived from the MFA for evidence of spatial autocorrelation using variograms in the *gstat* package in R (Pebesma 2004). The output is given in Figure 5.6, and shows the semivariance (half the squared difference in SEP score between pairs of households) on the y-axis and the Euclidean distance (between household pairs) on the x-axis. If spatial dependency (autocorrelation) existed in SEP, we would expect households that are closer together to have (on average) more similar scores, and a smaller semivariance value, which would tend to increase with distance. This does not seem to be the case, and suggests that spatial dependency is not a feature of these data.

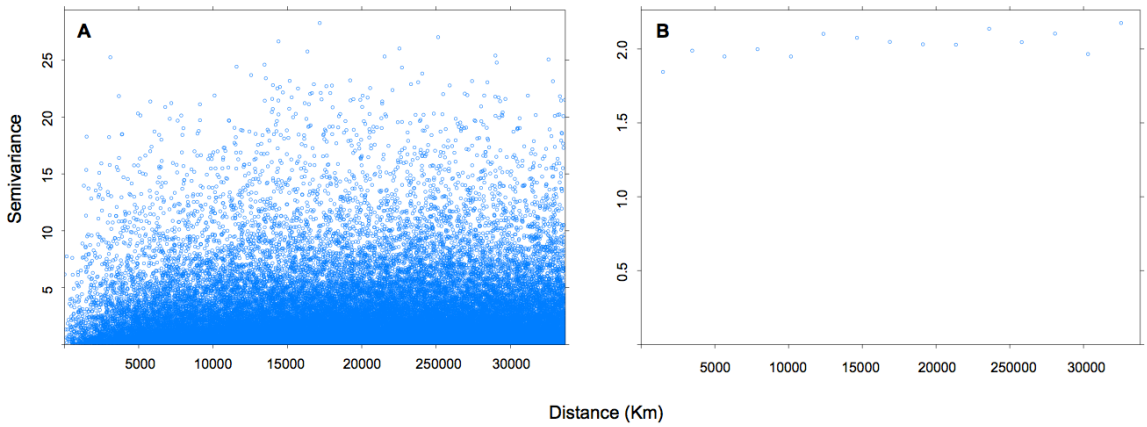


Figure 5.6. Variograms for the SEP index a) between pairs of households b) using average distances between households.

In the description of the conditions of household within the study area in the previous section (5.2.1) we identified that many of the features that were used to derive the SEP index (e.g. the presence of a latrine, building quality, education, source of an external income) were related to urban distance. It might therefore be expected that households in, or closer to,

urban areas might have a different SEP compared to more rural areas. In addition, households that are closer to main roads may benefit from improved infrastructure that may impact upon their SEP. We examined the effects of these two conditions using a simple multivariable linear regression. The method to define urban distance was described previously. We calculated the Euclidean distance between each household point and the nearest 'main' road (Kenya A, B and C roads which link provincial and/or district centres[4]) using the *Near* function in ArcMap 10.1.

Distance to an urban area and distance to main roads were not correlated (Spearman's rho = 0.05, p = 0.27). Prior to regression, both variables were scaled by subtracting the mean and dividing by the sample standard deviation (Schielzeth 2010). There was some evidence of a negative relationship between urban distance and SEP (β=-0.14, p=0.048) and much weaker evidence of a negative relationship with distance to main roads (β =-0.13, p=0.071), but this simple model explained an extremely small amount of the variation in SEP (R^2 = 0.02).

*****

5.2.3. Exploring the effect of household SEP on individual infectious disease risk

In chapter 4, we used multilevel logistic regression models to quantify the importance of household effects on explaining variation in individual risk of infection. We extend these models here, and examine how much of the between household-variation in risk SEP might explain.

Choice of outcomes

The PAZ study identified 21 different infectious agents in people in this community (described in chapter 4), however we examine here the effect of SEP on the most prevalent of these: viz. *P. falciparum*; *E. histolytica/dispar*; *A. lumbricoides (*hereafter, *Ascaris)*; hookworm; *T. trichiura (*hereafter, *Trichuris)* and TB infection (on the basis of a gamma-IFN assay, and this therefore includes latent TB infection, LTBI). Not only are these the infectious agents that affect the most people, and which have a range of transmission routes, but focusing on common outcomes ensures model stability whilst enabling the exploration of a range of effects. A sample size of 10 to 15 observations per predictor provides a rough rule

[4] Data generated from a 1:250,000 geo-referenced road map for Busia County downloaded from the Kenya Roads Board (http://www.krb.go.ke/road-maps.html).

of thumb for regression analysis (including logistic regression) (Harrell 2006; Burnham & Anderson 1998; Babyak 2004). The rarest infection, TB, with 150 cases, therefore allows us 10 to 15 degrees of freedom.

As described in the previous chapter, infection with *Taenia* spp. was also highly prevalent, and showed substantial clustering at the household level (suggesting household effects are important). However multilevel predictors of *Taenia* spp. infection are being explored by other authors (Wardrop et al *in prep*) and we do not include analysis of the parasite here.

Model building

We used the hierarchical approach to model building described by Victora et al. (1997). This involves defining a conceptual framework that describes the hierarchical (in terms of proximal and distal) relationships between the main predictors of interest and the outcome, and identifying covariates that may either confound or mediate the effect (Messer et al. 2008; Boccia et al. 2009). Confounders are related to the outcome and the exposure and are capable of distorting the observed relationship, but are not on the causal pathway between the two. Mediators are also related to both the outcome and the exposure but, importantly, *are* on the causal pathway. Hence, whilst confounders may bias the relationship, mediators will tend to attenuate the effect (Victora et al. 1997).

The hierarchical approach involves examining (and reporting) the effect of predictors with control for important confounders. Only then are mediators added to the model and their effects reported (to prevent the effect of the more distal determinant being diminished by the more proximal effects (Victora et al. 1994)).

We defined confounders and mediators of the effect of SEP on each infectious disease outcome using the simple conceptual frameworks in Figure 5.7. These effects are also discussed below.

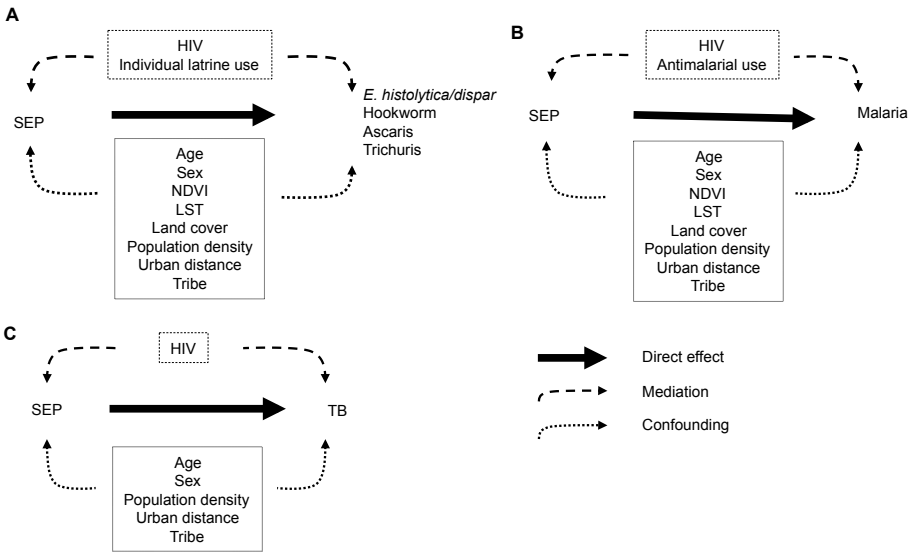


Figure 5.7. Conceptual pathway for multilevel analysis of the effect of SEP on the prevalent gastrointestinal parasites (A); Plasmodium *falciparum* malaria (B) and LTBI (C)

Confounders

Environmental effects

Environmental conditions define the broad geographic distribution of many infectious agents (particularly intestinal nematodes (Brooker et al 2006b) and vector borne-parasites (Kitron 1998) and may also impact upon transmission dynamics at small spatial scales through effects on the level of environmental contamination of free-living parasitic stages or vectors. The normalised difference vegetation index (NDVI), for example, a proxy measure of vegetation cover associated with both soil moisture and rainfall, has been shown to predict individual risk of infection with *Ascaris* (Saathoff et al. 2005), hookworm (Pullan et al. 2008) and malaria (de Oliveira et al. 2013; Hay et al. 1998). Similarly, variation in land surface temperature (LST) within small areas has been found to be associated with individual risk of hookworm (Riess et al. 2013) and *Ascaris* (Schüle et al. 2014) infection.

We used Fourier processed MODIS mean NDVI and maximum LST for the period 2001 to 2005 (data from Scharlemann et al. (2008)) as a possible confounder of the relationship between SEP and the parasitic infections (Figure 9). These data have a 1km spatial resolution, and the raster value at the location of each household was extracted.

Local land use may also be linked to household SEP, as well as having a direct impact on infectious disease risk, particularly for vector-borne and parasitic diseases (Krefis et al. 2011). Land cover has been defined at a very high spatial resolution (15 metres) for the study area using (ASTER) satellite imagery (Wardrop et al *in prep*). We were particularly interested in the area of grassland/crop land surrounding each household since this has been shown to have an effect on community-level poverty in western Kenya (Okwi et al. 2007). The proportion of an area with a radius of 200 metres surrounding each household that was classified as grassland and crops was extracted using ArcMap 10.1.

Population density

Population density is likely to be important for the transmission of many communicable diseases. To account for potential confounding with SEP at a 'local' spatial scale (which is the scale one might expect the main influence on infectious disease to occur), we generated a proxy of local population density by creating a georeferenced point for *all* identifiable human made structures (excluding structures that were obviously schools or churches) identifiable within a 200m radius of each sampled household using Google Earth satellite imagery. The resulting 'habitation points' were exported to ArcMap 10.1, where the *Kernel Density* function in the *Spatial Analysis* tool set was used to derive the density of points surrounding each household with a bandwidth of 200m and cell size of 20m. The resulting 'habitation density', assumed to represent local population density, was extracted at the household point.

Our bespoke population density estimate had no relationship whatsoever with that estimated by two publicly available data sets, namely Landscan 2008 (UT-Battelle, ONRI) (Spearman's rho = -0.03, p = 0.61) and Afripop (www.worldpop.org.uk) (Spearman's rho = 0.05, p = 0.29). However, these two commonly used sources of global population density (with different spatial resolutions) were very weakly correlated with each other when the value at households was compared (r = 0.09 p = 0.06). Clearly the procedure we adopted to define population density is subject to both omission and commission error: not all human structures identified as such will be occupied; natural features could be mistaken for man-made structures; dwellings may have been built or destroyed after the date the image was collected; multi-storied buildings may house many families. Despite this likely error, we felt the resulting output provided a better representation of local population density than the open-source alternatives (at the spatial scale of the study area).

It is important to note that the accuracy of Google Earth imagery can be highly variable, particularly in rural areas (Kamadjeu 2009; Chang et al. 2009). We assessed the positional accuracy of the images used to generate habitation points by comparing the Google earth imagery with ASTER satellite imagery (of 15 m resolution, and used for land classification described by Wardrop et al *in prep*). For this, a polyline feature for a random, but geographically representative, selection of linear features (roads, rivers, airstrips, field boundaries) observed in the ASTER imagery was generated in ArcMap 10.1 and exported to Google Earth for comparison with the imagery therein. In all cases, linear features identified using ASTER satellite imagery were within 25m of the location of the linear feature in Google Earth. All imagery in Google Earth was from 2010 or beyond, and was free of cloud cover.

Tribal affiliation

The area of high risk for multiple infectious diseases identified in chapter 4 (in the south and south east of the study area) was in a geographic area in which the Luo community are known to reside. We therefore included Luo tribal membership as a possible confounder of risk for each infection. Tribe was measured at the individual level, but we felt the group-level variable would provide better control for the contextual effect of SEP on infection risk. Household tribe (Luo/non-Luo) was defined on the basis of the reported ethnicity of at least 50% of people within a household (since we did not record the tribal affiliation of the household head, and this could not be determined from anonymised data).

These household level predictors, together with age and sex at the individual level, could be reasonably expected to predict both household SEP and individual infectious disease risk, but are unlikely to be on the causal pathway between the two.

Mediators

Relationships between individual HIV infection and each of the infectious diseases of interest have been described (Brown et al. 2006; Corbett et al. 2003; Van geertruyden & D'Alessandro 2007; Karp & Auwaerter 2007a), and HIV is also linked to SEP in sub-Saharan Africa (Hajizadeh et al. 2014). However, given that HIV could be on the causal pathway between SEP and infectious disease risk (SEP influences HIV risk; HIV infection influences risk for the other pathogens), we considered HIV positivity as a possible mediator in each case (Figure 5.7). For the gastrointestinal parasites, access to a latrine is likely to influence individual infection risk (Ziegelbauer et al. 2012). Whilst latrine ownership is

linked to SEP (and is present in our index), the behavioural factors that influence whether or not a latrine is used may also influence risk, and could therefore mediate its effect on risk for the gastrointestinal parasites (since not all who live in a household with a latrine will always use it).

The recent use of anti-malarials is likely to influence the likelihood that an individual has a detectable *P. falciparum* parasitaemia at the time of sampling, and may be used more often in households with a higher SEP. The use of anti-malarials could therefore mediate the relationship between SEP and individual malaria infection.

Variable structure and model selection

In general, we would expect SEP to be related linearly to each outcome (and that infectious disease risk would decrease or increase at a constant rate with increasing SEP). However, since our data set is reasonably large, and the chosen outcomes common, we allowed the data to decide on the shape of the relationship for the main predictor (SEP) as well as continuous covariates (included for their potential confounding effect) with the log odds of each infection.

The full model (i.e. containing SEP and hypothesised confounders, but not mediators) was fit with household as a random effect for each infectious outcome. Model improvement following the inclusion of administrative division as a second random effect (for a three-level logistic regression model) was assessed using the likelihood ratio test, and division retained if there was evidence of a significant improvement (at the 5% alpha level). The shape of the relationship between continuous variables and (the log odds of) each infectious outcome was initially explored using the quartile approach described by Hosmer et al. (2013). For this, a categorical variable with four levels was created on the basis of quartile cut-points. A multivariable model was then fit replacing the continuous predictor with its four-level categorical derivative, and the resulting coefficients plotted versus the mid-points of the upper three quartiles (with the lowest used as reference, and given a value of 0). Where the relationship was obviously linear, the continuous form was used. Where a logical parametric shape (e.g. quadratic) could be observed, made biological sense, and resulted in model improvement based on AIC, the variable was transformed. Where the relationship appeared more complicated, a restricted cubic spline (RCS) with either 3 or 4 knots was adopted. If the resulting fitted shape made biological sense, and resulted in lower AIC when compared to the model containing the linear (or squared) variable, the best fitting RCS (defined using default cut-points (Harrell 2006)) was used.

For ease of interpretation, we adopted a model selection procedure, and present the results of the most parsimonious model only. Model selection involved removing hypothesised confounders from the full model (note, this is still the model without the mediators) in a backwards step-wise procedure, starting with the least significant (on the basis of p-value from the Wald test). At each step, the benefit of retaining the variable was also assessed using the likelihood ratio test (with an alpha level of 5%). The effect of removal of a hypothesised confounder on the co-efficient of SEP was also examined. If removal resulted in more than 15% change, the covariate was retained regardless of p-value (Hosmer et al. 2013). The effect of SEP on each outcome in the final model, containing all significant predictors (at the 5% alpha level) or important confounders was recorded. Only then were mediators added. No variable selection procedure was used for the mediators.

Multilevel logistic regression models were fit using *lme4* in R (Bates et al. 2014). Restricted cubic splines were incorporated, where appropriate, via the *rms* package (Harrell 2006).

Model diagnostics

To ensure there was not excessive collinearity in the full model (containing SEP, confounders and hypothesised mediators), variance inflation factors (VIF) were calculated, and a cut-off of 2.5 used to indicate a potential issue (excluding those variables with restricted cubic splines), with a particular focus on examining collinearity with SEP.

In the previous chapter, we observed substantial spatial variation in household risk of infection for all of the infectious outcomes of interest. This spatial dependency is of substantive interest as it may suggest the existence of geographically varying contextual processes that influence infection risk and which we seek to explain. However, where the spatial autocorrelation is not completely explained by the covariates included in a regression analysis, residual spatial autocorrelation (RSA) violates the independence assumptions for (generalised) linear models and may affect estimated co-efficients and error probabilities (Kuhn & Dormann 2012; Kuhn 2007).

We derived the household-level residuals from each final model and examined these for the presence of RSA using Morans I. Where evidence of residual autocorrelation was observed, we attempted to better explain the autocorrelation by including appropriate predictors not already in the model but that could be hypothesised to better explain the spatial distribution for the infection under consideration (and we discuss what these might be in later sections).

Where this failed to account for RSA, the models were de-trended by including latitude and longitude (in an effort to account for broad spatial processes) and re-assessed for RSA.

The Moran's I statistic was calculated using the *ape* package in R (Paradis et al. 2004).

Proportional change in variance

In order to explore the proportion of between household variation in individual risk that SEP explains, we compared the full, final multivariable model for each outcome (containing all significant covariates and important confounders, but without mediators) with a reduced model without SEP. The proportional change in variance (PCV) was calculated as (Merlo et al. 2006):

$$\text{PCV} = \frac{V_A - V_B}{V_A} \times 100$$

Where V_A is the (household-level) variance of the reduced model without SEP and V_B the (household-level) variance of the full model. The PCV can therefore be considered somewhat analogous to R^2 and (broadly) indicates the explanatory power of the excluded variables (Hox 2010). The benefit of retaining SEP (in terms of model fit) was also assessed using a likelihood ratio test.

Results

Entamoeba histolytica/dispar; Plasmodium falciparum; TB

The results of the multilevel logistic regression analysis for *Entamoeba histolytica/dispar*, *P. falciparum* (malaria) and TB are presented in Table 5.2. There was strong evidence that infection risk for both *E. histolytica/dispar* and malaria was negatively associated with increasing SEP. Using the derived model coefficients to make predictions, and holding all other covariates at their reference value (in the case of the categorical variables) or their average (in the case of continuous covariates), we might expect that an individual living in the poorest household in our study area would have a probability of around 0.38 of being infected with *E. histolytica/dispar*, whilst it would be around 0.20 in the richest, or an effect size range of 0.18[5]. Similarly, an individual in the poorest household would be expected to have a probability of infection with *P. falciparum* of around 0.49, whilst it would be 0.21 in the richest, or an effect size range of 0.28.

[5] Effect sizes were calculated using the *plotLMER.fnc* in the *LanguageR* package in R, but this is essentially estimated as inv.logit(intercept + β_{SEP}* maximum SEP +...) – inv.logit(intercept + β_{SEP}* minimum SEP +…)

Despite these moderate effect sizes, SEP did not explain a great deal of the between-household variation for *E. histolytica/dispar* (5.5%), and only a modest amount for malaria (14.6%). Only age and sex were retained as covariates for *E. histolytica/dispar*, and appeared to have a minimal confounding effect on SEP. The non-linear relationship between age and *E. histolytica/dispar* infection is shown in Figure 5.8, and demonstrates a clear increase in infection risk into early adulthood, after which risk rapidly declines (and also shows the utility of using splines to model these complex non-linear relationships: the effect of age would have been missed entirely using a linear term, and is poorly represented with a quadratic term). Females had a higher risk of infection with *E. histolytica/dispar* than males.

Increasing local population density resulted in reduced risk of malaria, whilst increasing distance from an urban area increased risk, suggesting rurality is important. Neither effect was important as a counfounder of the relationship with SEP and malaria (i.e. when comparing the uni- and multivariable odds ratios in Table 5.2). The probability of malaria infection decreased linearly with age.

In contrast to both of these highly prevalent parasitic infections, the probability of LTBI appears to increase (linearly) with increasing SEP (Table 5.2). None of the other household-level covariates (population density, urban distance or household membership of the Luo tribe) appeared to be important as predictors of TB infection, or to have a confounding effect on SEP. The probability of LTBI increased linearly with age, but there was minimal evidence of a confounding effect on SEP. Infection with *Mycobacterium* spp. was relatively rare in our study area (on the basis of the final model, and holding SEP and age at their average values, the average probability of TB infection is estimated at around 0.03), but is predicted to be more than three times higher in individuals in the richest households compared to those in the poorest (a probability of infection of 0.09 in an individual in the richest vs. 0.027 in the poorest).

Table 5.2. The relationship between SEP and *E. histolytica/dispar*, *P. falciparum* and LTBI, with control for confounders and important covariates. The greyed cells represent variables that were not on the conceptual pathway or values that were not estimated; dashes represent variables that were not included in the final multivariable model.

	Entamoeba histolytica		***Plasmodium falciparum***		**LTBI**	
	uOR[1] (95% CI)	**mOR[2] (95% CI)**	**uOR[1] (95% CI)**	**mOR[2] (95% CI)**	**uOR[1] (95% CI)**	**mOR[2] (95% CI)**
SEP	0.85 (0.75-0.97)	0.84 (0.74-0.96)	0.75 (0.66-0.84)	0.77 (0.68-0.88)	1.26 (1.04-1.53)	1.26 (1.03-1.54)
NDVI	1.04 (0.91-1.19)	-	1.01 (0.90-1.13)	-		
LST	1.14 (0.98-1.33)	-	1.13 (1.00-1.29)	-		
Pop. density	0.94 (0.82-1.08)	-	0.79 (0.69-0.90)	0.84 (0.73-0.98)	1.11 (0.89-1.37)	-
Urban distance	0.97 (0.84-1.11)	-	1.13 (1.01-1.27)	1.17 (1.03-1.33)	0.96 (0.78-1.18)	-
Grass area	0.84 (0.37-1.88)	-	1.66 (0.82-3.38)	-		
Age	*See Figure 8*		0.73 (0.70-0.76)	0.73 (0.70-0.77)		1.14 (1.09-1.20)
Sex	0.75 (0.62-0.93)	0.76 (0.62-0.94)	1.21 (0.99-1.47)	-	1.13 (0.791.63)	-
Luo	0.91 (0.63-1.33)	-	1.09 (0.83-1.44)	-	1.45 (0.94-2.25)	-
Random effects						
σ^2_H (95% CI)[3]		0.58 (0.34-0.90)		0.32 (0.11-0.61)		0.65 (0.10-1.42)
σ^2_D (95% CI)[3]		0.066 (0-0.24)		0.023 (0 – 0.11)		-
PCV_1[a] (p-value)		5.5% (0.013)		14.6% (<0.001)		7.0% (0.02)

[1] Univariable odds ratio; [2] multivariable odds ratio; [3]estimated using profile methods (Hox 2010)
[a] Proportional change in variance (PCV) at the household level comparing full model with model without SEP, with p-value estimated on the basis of a likelihood ratio test

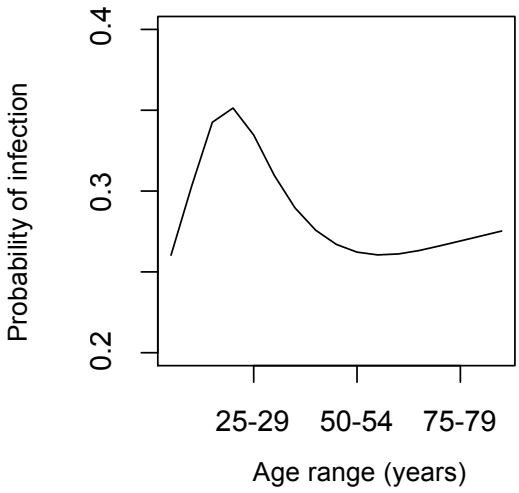


Figure 5.8. Predicted relationship between probability of infection with *E. histolytica/dispar* and age range using restricted cubic splines with 4 knots.

There was little evidence of a mediating effect by HIV infection on the relationship between any of these three infections and SEP (Table 5.3). Individual latrine use appeared to reduce the risk of *E. histolytica/dispar* infection, independent of SEP, but the evidence to support this effect was quite weak (i.e. the 95% CI includes 1, although the p-value of the likelihood ratio comparing nested models with and without individual latrine use was 0.03). As expected, those individuals who had recently taken anti-malarials (within the past month) had a lower probability of *P. falciparum* infection, but there was little evidence of mediation of SEP when comparing the odds ratio for SEP without mediation (Table 5.2) to that with mediation (Table 5.3) (i.e. 0.78 vs. 0.77).

There was no evidence of residual spatial autocorrelation or excessive collinearity for any of the final models (see Appendix 5.4 for model diagnostics).

Table 5.3. The effect of the hypothesised mediators on the relationship between SEP and *E. histolytica/dispar*, *P. falciparum* and LTBI. The greyed cells represent variables that were not in the conceptual pathway.

	Entamoeba histolytica		***Plasmodium falciparum***		**TB**	
	uOR[1] (95% CI)	**mOR[2] (95% CI)**	**uOR[1] (95% CI)**	**mOR[2] (95% CI)**	**uOR[1] (95% CI)**	**mOR[2] (95% CI)**
SEP	0.85 (0.75-0.97)	0.86[a] (0.76-0.99)	0.75 (0.66-0.84)	0.78[a] (0.68-0.89)	1.26 (1.04-1.53)	1.22[a] (1.0-1.50)
HIV	0.9 (0.57-1.43)	0.89[a] (0.55-1.43)	0.30 (0.17-0.55)	0.62[a] (0.34-1.16)	1.01 (0.45-2.23)	0.55[a] (0.23-1.27)
Latrine use	0.83 (0.67-1.04)	0.80[a] (0.63-1.01)				
Recent antimalarial			0.78 (0.56-1.08)	0.67[a] (0.47-0.96)		

[1] Univariable odds ratio; [2] multivariable odds ratio.
[a] includes control for all covariates shown in multivariable model for infection in Table 5.2.

The soil-transmitted helminths

The effect of SEP on *Ascaris*, *Trichuris* and hookworm is shown in Table 5.4. Only for hookworm was there any evidence of an effect of household SEP on the probability of individual infection, but this effect was quite substantial. On the basis of the final multivariable model, which also included NDVI and Luo tribal membership as household level predictors, we might expect that individuals living in the poorest household in the study area (and holding all other predictors constant) to have a probability of infection of around 0.54, whilst this would be just 0.1 in the richest households (an effect size range of 0.44). An individual's probability of hookworm infection increased with age, and this was best represented as a slight negative quadratic relationship. Individuals living in Luo households were at lower risk, whilst increasing local NDVI resulted in increased probability of individual infection, although this relationship was not linear (and included a negative quadratic term, suggesting increasing probability up to a plateau followed by a decline at higher values of NDVI). Males were more likely than females to be infected by hookworm.

The final model for *Trichuris,* which included as household-level predictors membership of the Luo tribe (increased individual risk), local population density (increased individual risk), and maximum land surface temperature (decreased individual risk with increasing LST), showed some evidence of residual spatial autocorrelation (RSA). In Chapter 4 we observed substantial overlap of the spatial distribution of *Trichuris* and HIV, and therefore before considering (somewhat complicated) spatial models to account for this residual spatial effect, we explored whether the inclusion of contextual effects of HIV infection at the household

level might explain this RSA. Since this is beyond the main focus of this chapter, the steps involved are discussed in more detail in Appendix 5.4. Briefly, we found that including the count of HIV positive people present in the household was an important predictor of *Trichuris* infection, and explained much of the RSA (or enough that the p-value for the Moran's I was above the 5% alpha level). The relationship between household HIV count and *Trichuris* infection is shown in Table 5.5, but it is perhaps useful to note that the odds ratio and the precision of the confidence interval for SEP is broadly the same as that shown in Table 5.4, suggesting the RSA has minimal impact upon either the co-efficient for SEP or its error. Also shown in Table 5.5. is the effect of individual HIV infection on the probability of *Trichuris* infection. Despite the spatial overlap of these infections, and the positive effect of HIV infection at the household level, individual HIV infection status was a negative predictor of the risk of *Trichuris* infection. This effect was observed both with and without the count of HIV positive individuals in the model (as evidenced by comparing the effect of individual HIV in the univariable and multivariable models in Table 5.5). This (somewhat unexpected) negative effect of individual HIV infection was also observed for hookworm, but not for *Ascaris*. Household HIV count did not impact upon hookworm or *Ascaris* (although the OR for the latter suggests a positive, but statistically insignificant relationship).

Individual latrine use had a significant effect on individual risk of hookworm infection, but it was necessary to control for SEP (and all other covariates in the multivariable model) in order to observe it (i.e. the effect was not significant at the univariable level). There was some indication of slight mediation of the effect of SEP on hookworm infection by the inclusion of HIV and individual latrine use (which changed from an OR 0.64 to 0.67).

The final model for *Ascaris* infection contained substantial RSA that was not reduced by including predictors of HIV infection at the household level. The inclusion of latitude and longitude, both of which were significantly associated with *Ascaris* infection (with individual infection appearing to be most likely in the south and east of the study area (data not shown)) was sufficient to eliminate the RSA. The outputs for this parasite in Table 5.4 and Table 5.5 therefore include control by latitude and longitude.

Table 5.4. The relationship between SEP and Hookworm, *Ascaris* and *Trichuris*, with control for confounders and important covariates.

	Hookworm		***Ascaris lumbricoides***		***Trichuris trichiura***	
	uOR[1] (95% CI)	**mOR[2] (95% CI)**	**uOR[1] (95% CI)**	**mOR[2,3] (95% CI)**	**uOR[1] (95% CI)**	**mOR[2] (95% CI)**
SEP	0.69 (0.59-0.81)	0.64 (0.55-0.76)	0.82 (0.58-1.16)	0.85 (0.60-1.21)	1.03 (0.82-1.30)	0.95 (0.76-1.20)
NDVI	1.17 (0.97-1.41)	1.22 (1.01-1.48)	1.27 (0.87-1.84)	-	0.96 (0.77-1.20)	-
$NDVI^2$	0.84 (0.74-0.95)	0.83 (0.73-0.94)				
Max LST	1.12 (0.9- 1.36)	-	0.89 (0.63-1.24)	-	0.79 (0.62-1.01)	0.8 (0.62-1.02)
Grass area		-		-		-
Pop. density	0.81 (0.67-0.97)	-	1.01 (0.7-1.47)	-	1.39 (1.13-1.70)	1.39 (1.12-1.72)
Urban distance	1.19 (1.0-1.41)	-	1.02 (0.72-1.44)	-	1.10 (0.88-1.38)	-
Age	1.27 (1.15-1.40)	1.31 (1.18-1.45)	0.84 (0.78-0.90)	0.84 (0.78-0.90)	0.92 (0.88-0.96)	0.92 (0.88-0.97)
Age^2	0.99 (0.98-1.0)	0.99 (0.98-1.00)				
Sex	1.31 (1.06-1.61)	1.43 (1.15-1.77)	0.90 (0.6-1.37)	-	0.65 (0.47-0.91)	0.65 (0.47-0.91)
Luo	0.59 (0.37-0.96)	0.58 (0.36-0.92)	2.97 (1.41-6.25)	-	3.23 (1.53-6.80)	3.54 (1.74-7.24)
Random effects						
σ^2_H (95% CI)[4]		1.31 (0.91-1.83)		4.0 (2.42-7.02)		1.18 (0.64-2.0)
σ^2_D (95% CI)[4]		0.09 (0-0.42)		-		0.8 (0.18-2.96)
PCV_1[a] (*p*-value)		13.6% (<0.001)		3.1% (0.27)		-0.3% (0.67)

[1] Univariable odds ratio; [2] multivariable odds ratio; [3] de-trended by inclusion of latitude and longitude to account for residual spatial autocorrelation; [4]estimated using profile methods

[a] Proportional change in variance (PCV) at household level comparing full model with model without SEP, with p-value estimated on the basis of a likelihood ratio test

Table 5.5. The effect of the hypothesised mediators on the relationship between SEP and the STH.

	Hookworm		***Ascaris lumbricoides***		***Trichuris trichuria***	
	uOR[1] (95% CI)	**mOR[2] (95% CI)**	**uOR[1] (95% CI)**	**mOR[2,3] (95% CI)**	**uOR[1] (95% CI)**	**mOR[2] (95% CI)**
SEP	0.69 (0.59-0.81)	0.67[a] (0.57-0.79)	0.82 (0.58-1.16)	0.86 [a] (0.60-1.23)	1.03 (0.82-1.30)	0.93 [a] (0.73-1.19)
HIV	0.31 (0.17-0.56)	0.23[a] (0.13-0.43)	0.43 (0.15-1.23)	0.69 [a] (0.24-1.96)	0.33 (0.15-0.70)	0.22 [a] (0.10-0.52)
Individual latrine use	0.81 (0.64-1.03)	0.69 [a] (0.53-0.89)	0.81 (0.50-1.30)	0.93 [a] (0.56-1.54)	0.76 (0.53-1.09)	0.79 [a] (0.54-1.16)
HIV count	0.70 (0.55-0.91)	1.03 [a] (0.78-1.35)	1.36 (0.85-2.19)	1.52 [a] (0.93-2.49)	1.46 (1.11-1.92)	1.53 [a] (1.13-2.08)

[1] Univariable odds ratio; [2] Multivariable odds ratio; [3] de-trended by inclusion of latitude and longitude to account for residual spatial autocorrelation. [a] includes control for all covariates shown in multivariable model for infection in Table.

Further exploration of the effect of SEP on infection risk

The relationship between SEP and access to medical services

A possible explanation for the observed elevated risk of *E. histolytica*, hookworm and malaria in individuals living in poorer households might be as a result of reduced access to medical services. To provide further evidence for this, we looked at the relationship between household SEP and recent individual use (within the past month) of antibiotics, anti-malarials and anti-inflammatories (which could be considered as proxies of a more general access to medical services). These outcomes were reported reasonably commonly within the study area (8.9% (95% CI 7.2-10.5), 10% (8.4-11.7), 29.1% (26.7-31.5), respectively) whilst the (reported) use of anthelmintics (0.9% (0.4-1.5) and anti-retrovirals (0.9 (0.4-1.4)) were exceedingly rare.

For each outcome (recent use of antibiotics, antimalarials or anti-inflammatories), we fitted multilevel logistic regression models following the general scheme described above (i.e. assess the benefit of including division as a second random effect in addition to household; assess the functional form of all linear predictors, and choose the most appropriate shape; fit the full model with SEP and all covariates and use backward step-wise selection to identify the most parsimonious model).

Only distance to an urban area, age and sex were considered as confounders of the relationship between SEP and each outcome and we did not consider the effect of mediators.

The results are presented in Table 5.6, and suggest that increasing household SEP increases the likelihood that an individual uses each of these drugs. The estimated probability that an individual in the poorest household has used antibiotics in the past month would be around 0.03 whilst it would be around 0.15 (holding sex at reference and the average for the continuous predictors) in the richest household (an effect size range of 0.1^2). Similarly, an individual in the poorest household would be expected to have a probability of around 0.2 of using anti-inflammatories in the past month (and these are expected to be predominantly paracetomol and aspirin), whilst it is estimated to be around 0.4 in the richest household. The relationship with anti-malarials is almost linear, but the inclusion of a restricted cubic spline with 3 knots provided the best fit (Figure 5.9) and suggests that there is an initial decrease in the use of antimalarial from the poorest to the slightly less poor households, but then follows the expected pattern of increased usage as household SEP increases.

There was a reasonably high amount of between-household variation in the probability of individual use for each outcome[6]. Socioeconomic position explained quite small amounts of this variation for all medicines, being most important for recent antibiotic use (explaining 15% of between household variation in risk).

There was no evidence of residual spatial autocorrelation in any of the models (see Appendix 3.4).

[6] Household level MORs, the concept of which was introduced in Chapter 4, and which provide a measure of between-group variation, and VPCs, which provides a measure of within-group clustering were (on the basis of the null model without covariates but with household and division as random effects): antibiotics: MOR = 3.1, VPC= 30.3%; anti-inflammatories: MOR = 2.6, VPC= 23.6%; anti-malarials: MOR = 2.9, VPC = 27.9%).

Table 5.6. The relationship between SEP and recent use of medicines with control for confounders and important covariates. The greyed cells represent variables that were not in the conceptual pathway or values that were not estimated; dashes represent variables that were not included in the final multivariable model.

	Antibiotics		**Anti-inflammatories**		**Antimalarials**	
	uOR[1] (95% CI)	**mOR[2] (95% CI)**	**uOR[1] (95% CI)**	**mOR[2] (95% CI)**	**uOR[1] (95% CI)**	**mOR[2] (95% CI)**
SEP	1.36 (1.13-1.64)	1.37 (1.14-1.66)	1.23 (1.09-1.39)	1.22 (1.08-1.38)	*See Figure 5.9*	
Urban distance	0.82 (0.66-1.02)	-	0.83 (0.73-0.96)	0.85 (0.74-0.97)	0.84 (0.70-1.02)	-
Age	1.04 (1.00-1.08)	-	1.02 (1.00-1.05)	-	0.80 (0.70-0.92)	0.78 (0.68-0.89)
Age^2					1.01 (1.00-1.02)	1.02 (1.01-1.03)
Sex	0.50 (0.36-0.71)	0.50 (0.35-0.70)	0.75 (0.61-0.92)	0.74 (0.61-0.91)	0.70 (0.52-0.95)	0.66 (0.48-0.89)
Random effects						
σ^2_H (95% CI)		0.68 (0.26-1.32)		0.39 (0.20-0.66)		0.88 (0.44-1.52)
σ^2_D (95% CI)		0.59 (0.16-1.91)		0.56 (0.26-1.35)		0.38 (0.14-1.03)
PCV_1[a]		15.0% (0.001)		7.4% (0.002)		3.1% (0.01)

[1] Univariable odds ratio; [2] multivariable odds ratio.
[a] Proportional change in variance (PCV) comparing full model with model without SEP

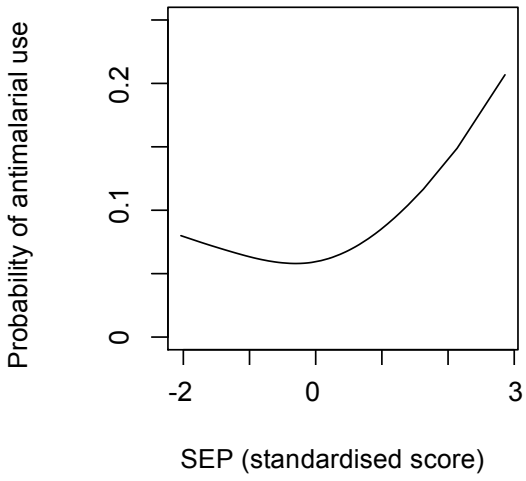


Figure 5.9. The predicted (non-linear) relationship between antimalarial use and SEP

5.3. Discussion

With the increasing availability of reasonably straightforward techniques to define SEP (Filmer & Pritchett 2001; Howe et al. 2012), epidemiological studies are increasingly exploring social gradients in infectious disease risk in low income countries. To date, these studies have focussed on single infectious outcomes, particularly HIV (Hajizadeh et al. 2014), TB (Oxlade & Murray 2012; Boccia et al. 2011; Boccia et al. 2009), malaria (Mmbando et al. 2011; de Castro & Fisher 2012; Woyessa et al. 2013; West et al. 2013; Pullan et al. 2010a; Ernst et al. 2009; Somi et al. 2007; Brooker et al. 2004b), and intestinal helminths (Schüle et al. 2014; Riess et al. 2013; Raso et al. 2004; Raso et al. 2006a). A smaller number of studies have included an assessment of the impact of SEP on helminth-malaria co-infection (Brooker et al. 2012; Righetti et al. 2012; Florey et al. 2012; Tshikuka et al. 1996; Raso et al. 2005), and gastrointestinal parasitism with a range of species (Hughes et al. 2004; Halpenny et al. 2013; Schmidlin et al. 2013; Matthys et al. 2011; Nundy et al. 2011; Knopp et al. 2010; Pullan et al. 2008; Raso et al. 2006b). Our study extends the assessment of the impact of SEP on multiple infectious diseases within a single community, and examines this relationship with a wide a range of endemic pathogens with diverse transmission routes.

The community under study is characterised by high levels of poverty. Busia district, which covers the majority of the study area, is the 10th poorest out of the 70 districts in Kenya (Open Kenya 2014), a country which ranks 147th out of 187 on the human development index (UNDP 2014). The apparent homogeneity of households in the area, in which the majority of people live in dwellings constructed from local materials (although most compounds also included at least one 'improved' building) and have limited access to 'western' levels of sanitation, piped water or mains electricity, could lead outsiders to conclude that such populations are relatively homogeneous in terms of their socioeconomic status (Bull 2009; Gwatkin 2003; Schellenberg et al. 2003). However, we identified a clear socioeconomic gradient in this community, which impacts both on infectious disease risk and access to medical care.

Our study supports the general trend for a positive effect of SEP on health (Marmot 2005), and shows reduced individual risk of infection with hookworm, malaria and *E. histolytica/dispar* (the three most prevalent infections in our study area) with increasing household wealth. However, we could find no evidence of an effect of SEP on *Ascaris* or *Trichuris*, and we observed a negative gradient for infection with *Mycobacterium* spp.

Individuals living in households with higher SEP were also more likely to report using antibiotics, anti-inflammatories and antimalarials in the past 4 weeks. In Tanzania, it was suggested that children in the lowest wealth quintiles did not have a different risk of disease, but were less likely to seek and receive health care and were therefore disproportionately burdened (Schellenberg et al. 2003). This might also contribute to the effect we observe in our own study area: people have lower levels of infection because they have greater access to medical care. Although the reporting period was quite short, very few people reported using anthelmintics, and therefore this does not seem to be an explanation in the case of hookworm. The (worryingly) high level of antibiotic use reported in this community could impact upon *Entamoeba* prevalence, however only particular classes of antibiotic will be effective against this organism (Mandell et al. 2010): we do not have data on the classes of antibiotics used. It is perhaps also useful to note that several authors have found that richer individuals, including those in rural sub-Saharan Africa (Raso et al. 2005; Zere & McIntyre 2003), are more likely to report ill health than poorer individuals in the same population. It might therefore be the case that richer individuals perceive that they are sicker and seek medical treatment more often (and of course also have the wherewithal to do so) despite not necessarily having the same levels of disease (and rates of infection) as poorer individuals.

Although there is some debate about whether malaria is an infectious disease of poverty (at least at the micro-level) (Worrall et al. 2005), several studies have shown that the richest individuals are least likely to have *P. falciparum* parasitaemia (Schellenberg et al. 2008; Njau et al. 2006; Somi et al. 2007), are more likely to use anti-malarials (Njau et al. 2006; Filmer 2005) and to seek them more rapidly (Chukwuocha et al. 2014). We observe the same negative association between parasitaemia and increasing SEP, as well as a positive association (albeit a moderately non-linear one) with anti-malarial use. The increased use of antimalarials by individuals living in richer households may explain some of the observed effect on reduced risk of infection (Schellenberg et al. 2003), although we observed a minimal mediating effect of antimalarial use, suggesting that the observed effect is not entirely due to increased treatment (although reports of use in the last 4 weeks may not entirely capture this). Other factors, such as sleeping under a bed net (and whilst most households reported owning bed nets, this does not necessarily mean that they are used (Githinji et al. 2010)), or housing quality (Lindsay et al. 2003), may be more important as mediators.

The negative effect of SEP on hookworm infection was quite large, and larger than that for any other infection. The relationship between poverty and hookworm is well known (Hotez

2008), and indeed the large effect observed could almost be considered to validate our index of SEP (although the PCV, and therefore (broadly speaking) the explanatory power of our derived index was quite low). As we described in detail in section 3.2.2, several factors may link SEP to infectious disease risk. In the case of hookworm, these are likely to include the availability and use of household sanitation and the resulting level of parasite contamination in the domestic and peri-domestic environment, as well as the level of skin contact with soil (which contains infective larvae). The effect of SEP on these exposures may be mediated through factors such as agricultural occupation (Hotez 2008), shoe wearing (Phiri et al. 2000), education (Asaolu & Ofoezie 2003; Pullan et al. 2010b) and household building materials (Pullan et al. 2011b). Nutritional status is also well known to impact upon susceptibility to helminth infection (Zaralis et al. 2009; Houdijk et al. 2009).

Given the observed relationship with hookworm, it is surprising that we did not observe a similar effect for *Trichuris* and *Ascaris*. These parasites are also transmitted via the soil and have broadly similar ecologies (although, and as we describe, and as is well known (Bethony et al. 2006), there are important differences in age groups affected). Other workers reporting results from sub-Saharan Africa have found the expected negative association with increasing SEP for *Ascaris* (Schüle et al. 2014; Kightlinger et al. 1998), although there seems to be less evidence for *Trichuris*. Diagnostic tests for the soil-transmitted helminths have notoriously low sensitivity, particularly on the basis of a single faecal examination (as used in this study) (reviewed in Appendix A.4.2), which might negatively bias the observed relationships. However, in such a case we would also expect a similar effect to occur for hookworm, and indeed that the poorest individuals might be those with the most intense infections for all three of the STH (since exposure would be expected to be greatest amongst the poorest) and therefore the bias should actually be positive (i.e. individuals with the most intense infections shed eggs in highest concentration and therefore are easiest to detect on microscopy). We explore the issues related to diagnostic test error in exploring risk factors for the STH, and the factors that may confound detection, in the next chapter. It is also worth pointing out that the sign of co-efficient for SEP is in the expected direction for both parasites (and indicates the expected decrease in risk with increasing SEP, particularly for *Ascaris*). Therefore this may be an issue with power rather than a lack of effect: in Chapter 2 we observed that there was substantial clustering of both parasites within households, as is widely reported in the literature. Even with our large sample size, such clustering might be expected to reduce the effective sample size when household membership is taken into account (via the random effect in the multilevel models), and may therefore affect our ability to observe a significant effect.

The finding that both hookworm and *Trichuris* are negatively associated with HIV is also unexpected. This is particularly interesting (and surprising) in the light of the positive relationship observed between individual *Trichuris* infection and the count of HIV positive individuals at the household level, as well as the extensive overlap in the spatial distribution of household risk for both infections observed in Chapter 4. Our main interest here was examining the potential mediating effect of HIV on SEP, of which there was little evidence, but we explore these interesting findings in more depth in the next chapter.

A positive gradient for infection with *E. histolytica/dispar* has previously been observed in several countries (Duc et al. 2011; Schmidlin et al. 2013). These parasites are transmitted via the faecal-oral route, and the use of contaminated water sources in the poorest households, or inadequate hygiene practices, are likely to be particularly important (Jensen et al. 2002; Cairncross et al. 2010)

Although few studies have explored the relationship between SEP and latent infection with *Mycobacterium* spp. (LTBI), Boccia et al. (2009) report that LTBI (on the basis of a gamma-IFN based assay, as used here) was most prevalent in the wealthiest individuals in a cross-sectional survey in Zambia. A similar effect has been reported for clinical (active) TB in a single population studied at different time points in Malawi (Odone et al. 2013; Glynn et al. 2000). These findings suggest TB infection might not follow the standard positive social gradient in health, although the economic costs of clinical TB may push individuals and households into poverty (Barter et al. 2012). Our results provide further support to the conclusion that there is a negative gradient for TB risk, although SEP explained only a relatively small amount of the between household variation. Infection with HIV has been repeatedly shown to be most prevalent in individuals with higher socioeconomic status (Hajizadeh et al. 2014), and TB-HIV co-infection might therefore explain the observed effect. However, in our population HIV did not appear to have a mediating (or confounding, although that was not our hypothesised mechanism) effect on the relationship between HIV and TB. The relationship was also not confounded by distance to urban areas or population density. The evidence that a negative gradient exists for infection with TB is perhaps made even more compelling given that we observe the expected positive gradient for infections such as hookworm, malaria and *E. histolytica/dispar* in the same community. We did not explore mediation to a great extent for this infection, but potential factors that link SEP to increased *Mycobacterium* spp. infection might include smoking and alcohol consumption (Oxlade & Murray 2012) and time spent in confined, poorly ventilated places such as bars, churches or public transport (Odone et al. 2013).

An alternative explanation for the observed relationship is confounding of the gamma-IFN assay through immunomodulation by helminth infection. As we described in the introductory chapter, helminth infection may be associated with a Th2 shift, and subsequent dampening of the Th1 immune response which may result in reduced production of gamma-IFN (a Th1 cytokine) in response to *Mycobacterium* antigens (Thomas et al. 2010). A systematic reduction in diagnostic sensitivity in the poorest individuals could therefore potentially contribute to the effect observed. Quinnell et al. (2004), for example, found some evidence of a suppressive effect of both hookworm and malaria infection (both of which we identified as being more common in poorer individuals) on production of gamma-IFN in response to PPD (i.e. tuberculin antigen). This potential confounding effect should be explored in more detail, and our data set (describing individual co-infection) will allow us to do that.

To our knowledge, ours is one of very few studies to use multiple factor analysis for the derivation of household socioeconomic score. In order to validate our use of MFA, we describe a novel application of the analytic hierarchy procedure in order to assign household SEP. This subjective approach had very high levels of agreement with the index define by MFA. The method (described in Appendix 5.3) needs further work, but we feel provides a reasonably simple and intuitive addition to current approaches used to define SEP, or at least to validate outputs derived from multivariate approaches.

Whilst SEP was a significant predictor for several infections, in no case did it explain a great deal of the between household variation in individual risk. The fact that it explains some of the variation is an outcome of substantive interest, but future work could examine whether alternative indicators of SEP might explain more of the this variation.

The vast majority of studies that explore the effect of a social gradient on infectious outcomes via multivariate methods (typically PCA) have categorised the derived index into quantiles. This follows Filmer & Pritchett (2001) who were early proponents of the use of PCA to derive household SEP. However, flexible parametric approaches to model non-linear relationships (including restricted cubic splines as used here, and fractional polynomials as used elsewhere (Riess et al. 2013)) are becoming increasingly available and straightforward to use. Whilst exploring inequality between socioeconomic strata may be useful in some cases, particularly for the targeting of interventions, the categorisation of continuous variables in regression analyses, which inevitably results in a loss of power and precision, is perhaps becoming increasingly difficult to justify (Bennette et al 2012).

Although we included the potential mediation effect of HIV co-infection, we did not explore co-infection or co-morbidities in this study, which would be an interesting next step. It

would seem likely that the poorest individuals are those in which multiple infections with malaria, hookworm and *Entamoeba histolytica/dispar* occur most frequently.

Since this is a cross-sectional study, we can of course say little about the direction of causation. We have tended to assume that poorer individuals have a different risk of infection than richer individuals, rather than people have different SEP because of infection. Ill-health in a sector that relies heavily on manual labour could result in decreased agricultural productivity, which may perpetuate poverty and intensify or sustain illness (Fox et al 2004, Audibert 2010).

Summary of main findings

1. A social gradient exists in this poor farming community that impacts upon individual infection with multiple infectious agents;
2. The poorest individuals are least likely to report using medical treatments in the past 4 weeks;
3. Individual infection with latent TB appears to increase as household-level SEP increases, but further work is needed to explore potential confounding of the observed effect by helminth infection.
4. There was minimal confounding of the effect of SEP on infectious outcomes by a range of household level predictors, and minimal evidence for mediation by HIV infection.

Chapter 6

A household-level analysis of soil-transmitted helminth risk

6.1. Introduction

Infectious disease epidemiology has tended to focus inference at the individual-level (Susser 1998; Rezaeian et al. 2007; Diez-Roux & Aiello 2005). Very often, community-based studies sample individuals from within a group (typically those people living in the same household) and examine how individual- as well as group-level effects influence individual risk of infection. The widespread use of multilevel models has (generally) allowed these cross-level effects to be modelled without bias (Diez-Roux & Aiello 2005). In some circumstances, it may also be useful to target inference at the group from which the individuals come (Susser 1994; Susser & Susser 1996; Greenland 2001). Defining a group of individuals as a 'patch' in which an infectious agent can be present or absent, for example, could be used for disease modelling and spatial analysis (Hanski 1998; Elith & Leathwick 2009; Stevens & Pfeiffer 2011) and for the identification of factors that may allow surveillance and control to be targeted at the group-level (Stärk et al. 2006).

As we described in chapter 4, defining group infection status as the presence or absence of infection in samples from group members may be complicated by poor test performance, or because only a proportion of all possible individuals within a group are sampled. A lack of detection of infection therefore does not mean the infection is absent. Without accounting for detection error, between-group prevalence estimates will tend to be biased (Branscum et al. 2004; McV Messam et al. 2008), relationships between predictors and a group-level response may be over- or under-estimated (Gu & Swihart 2004; Tyre et al. 2003), and variance estimates may be too narrow (MacKenzie et al. 2006). More subtly, within and between-group variability in the response to particular conditions may also bias inference at the group-level when these are unaccounted for (Kery & Schaub 2012).

In this chapter, we describe the application of zero-inflated binomial (ZIB) regression to estimate the probability that a household is 'infected' with, or occupied by, a pathogen when the pathogen is unobserved in samples collected from individuals living within it. These models, which couple the presence of an outcome at the group-level with the observation process in at the individual level (Kery & Schaub 2012), allow inference on the probability

of group-level infection, whilst controlling for the probability of detection in sampled individuals.

We applied ZIB, or 'site occupancy' models as they have become known in ecology (MacKenzie et al. 2002) to household-level infection with three environmentally transmitted helminths, *viz.* hookworm (due to *Ancylostoma duodenale* or *Necator americanus*), *Ascaris lumbricoides* and *Trichuris trichiura*. These soil-transmitted helminths (STH) cause anaemia and protein malnutrition (Rajagopal et al. 2014), and have been linked with growth retardation and cognitive impairment when acquired during childhood (Brooker 2010). The STH infect hundreds of millions of people throughout the developing world (Pullan et al. 2014b) and whilst much has been done to explore factors that influence individual risk of infection within endemic populations (Brooker et al 2006b; Bethony et al. 2006; Scott 2008), comparatively few studies have used household-level infection as a focus for analysis (but see Olsen et al. 2001; Redlinger et al. 2002; Trönnberg et al. 2010).

We used zero-inflated binomial models to try to answer the following questions:
1. What is the *true* prevalence of households infected, or occupied, by each STH species in an endemic area in western Kenya?
2. What are the factors that influence whether or not a household is occupied by each species?
3. What are the factors that influence whether or not we detect each species in an individual given that they live in an occupied household?

6.1.1. Background on the soil transmitted helminths

The STH (along with several other helminth species) are unique amongst the infectious agents of man in that they are non-replicating within the human host (Warren 1981). Instead, STH infected individuals shed oocytes in their faeces, which hatch and moult into infective larvae (hookworm) or embryonate and become infective eggs (*Trichuris* and *Ascaris*) in soil under appropriate conditions. It is exposure to infective larvae or eggs that results in infection or reinfection, and each individual adult STH present within the human host is the result of a single exposure. Two distinct populations can therefore be considered to exist: juveniles and adult worms in the human gastrointestinal tract, and eggs and larvae in the environment. In areas without universal access to high standards of sanitation, the population in the environment can be many orders of magnitude greater than that within the human host: female *Ascaris* worms can produce up to 200,000 eggs a day (Crompton 2001);

Trichuris up to 20,000 per day (Bundy & Cooper 1989), and adult female hookworms up to 10,000 eggs per day (Hotez 2005).

Environmental, social, behavioural, nutritional and genetic factors are likely to play a major role in determining exposure and/or susceptibility to infection, and therefore worm burdens (intensity), which tend to be highly aggregated in individuals (Anderson & May 1992). The diagnostic tests for the STH are generally based on the identification of eggs in faeces by microscopy (as was the case in this study: described in chapter 3). The probability of detection is therefore linked to the concentration of eggs within a faecal sample (Krauth et al. 2012), which is determined (albeit not perfectly linearly, and with some density dependent effects (Anderson & Schad 1985)) by the number of adult worms present in the small intestine. Diagnostic sensitivity may therefore be confounded by intensity of infection, and the factors that influence it.

The household provides a natural focus for the study of the STH (Otto et al. 1931), and in the case of *Ascaris* and *Trichuris*, is thought to be the 'domain' in which most transmission occurs (Cairncross et al. 1996; Bundy & Medley 1992). This is evidenced by studies, including our own in chapter 4, which have described strong clustering of STH infection within households in endemic areas (Walker et al. 2011; Pullan et al. 2010b; Brooker, Alexander, et al. 2006a; Forrester et al. 1988; Haswell-Elkins et al. 1989; Criscione et al. 2010). Particular families tend to be predisposed to high intensity infection and reinfection following treatment (Forrester et al. 1990; Chan et al. 1994b), which is likely to be due to a mixture of shared genetic and environmental effects (Ellis & McManus 2009; Pullan et al. 2010b; Pullan et al. 2008; Walker et al. 2011).

6.1.2. Background on the statistical method

'Site occupancy' models are based around two conditional Bernoulli processes (MacKenzie et al. 2002; MacKenzie et al. 2006):

$z_i \sim Bernoulli(\psi)$ True occupancy state

$y_{ij} \sim Bernoulli(z_i \times p_{ij})$ Sampling process yielding observations

The true occurrence of a species within site *i* can be considered a Bernoulli random process determined by parameter ψ, the probability that the site is occupied by the species of interest. The identification of a species during survey *j* of each site can also be considered to be a Bernoulli random process, in which the probability of success (detection) is the product of the true presence or absence of the species at the site (z_i) and the detection probability

(p_{ij}). Detection probability is therefore conditional on occupancy: if the species is not present at the site, we do not expect to detect it in any of the surveys (and therefore these models assume perfect specificity). Both the probability of occupancy and detection are estimable on the basis of repeat surveys of the site in either time or space (MacKenzie et al. 2002). Repeat visits, which must be more than one for some, but not necessarily all sites (Kery et al. 2010), allow the specification of a set of "detection histories" which, via multinomial maximum likelihood procedures, allow the estimation of the parameter of interest, ψ, the probability of occupancy, as well as p, the probability of detection given occupancy.

Importantly, it is possible to allow both ψ and p to vary on the basis of covariates. If we define 'occupancy' of a household as the presence of at least one infected individual, or the presence of at least one parasitic life stage in the household environment, covariates on STH occupancy might include factors related to the environmental or social conditions of the household, as well as those of the local community (and therefore community level prevalence). The probability that the parasite will be detected in sampled individuals given that they live in an occupied household will be a mixture of the probability that the sampled individual is infected *and* the probability that we detect the infection. Hence, possible covariates on detection will include conditions that influence the dynamics of transmission within a parasite occupied household and its peri-domestic environment, as well as factors related to the shedding of parasite eggs in infected individuals in sufficient quantities to allow detection (e.g. the intensity of infection and fecundity of resident parasites).

Site occupancy models can be considered as a hierarchical, coupled logistic regression analysis: one logistic regression describes occurrence of the species at the site; the other describes detection in each survey given that the species occurs in the site (Kery & Schaub 2012):

$$Logit(\psi_i) = \beta_0 + x_i B_1 \quad \text{and} \quad Logit(p_i) = \gamma_0 + x_i \gamma_1$$

where β_0, β_1, γ_0 and γ_1 are the unknown parameters that we aim to estimate. This hierarchical, conditional structure allows us to test competing hypotheses about the system in a reasonably straightforward way, and to identify factors that influence occupancy whilst controlling for detection, as well as those factors that influence detection given occupancy (Kery et al. 2010).

6.2. Methods

We used zero-inflated binomial ('site occupancy') models to explore factors that influence household infection for each STH species. In our application of these models, sites were households and 'occupancy' was household level infection with each STH species (treated individually), defined on the basis of the probability of detection of parasites in a single faecal sample collected from people living within.

6.2.1. Defining household 'infection'

A major assumption of site occupancy models is that a species is either present or absent within the larger primary sampling site during surveys of all subplots, and that there is a non-zero probability of detecting the species at all subplots (in our case, sampled people) given that the primary site is occupied (MacKenzie et al. 2006). We must therefore consider parasite occupancy as a truly household-level process, and that it is possible for the parasite to be present in the household environment but not necessarily in any of the individuals we sample. Household members tested for the presence of each parasite could therefore be considered as 'samplers' of the household environment, and we model the status of a household containing individuals rather than the status of individuals within a household. This assumption seems reasonable for environmentally transmitted parasitic infections, and particularly for the STH, where within-household transmission is known to be important, and for which environmental occupancy can be considered a state that is fundamental to transmission.

By defining the household as the *domain in which domestic transmission occurs* (Cairncross et al. 1996), household 'infection' can be influenced by the infection status of all those people we sample, as well as all those people we don't, the neighbour who defecates in the household's fields, the student who comes home at weekends, and the husband who works in Nairobi. This broad definition also allows us to meet the parametric (binomial) assumptions of the model, since the population of parasites within a household can be very large, even though the number of people we sample is small and finite.

6.2.2. Model specification

We explored a range of factors that are expected to influence both household occupancy and the probability that we detect infection in individuals given that they live in an occupied

household. The selected covariates are shown in Table 6.1, and the rationale for selection and source of data used to represent them is described below.

Covariates on probability of occupancy

i) Local population density

One might expect that a greater density of people surrounding a household would lead to higher levels of environmental contamination with parasitic life stages, and enhanced risk of infection at both the individual and household-level, and therefore in the wider community. However, literature-based reports of the magnitude and direction of the effect of population density on individual STH risk are conflicting (Riess et al. 2013; Ndenecho et al. 2002; Koroma et al. 2010). One explanation for the absence of an obvious effect might be that the relationship is non-linear: risk could increase until a cut-off at which point an area moves from being predominantly 'rural' to predominantly 'urban', with a resulting reduction in high risk behaviours (e.g. walking barefoot; agricultural work and contact with soil) or suitable habitat for the development of the free-living STH stages (Phiri et al. 2000; Brooker et al. 2006a; Pullan et al. 2010b).

We used the bespoke population density surface described in Chapter 5 to estimate the probability of household infection (occupancy), and included this together with a quadratic term to account for possible non-linearity.

ii) Homestead socioeconomic position

A relationship between socioeconomic position (SEP) and STH risk is very widely reported (Hotez 2008; Raso et al. 2006a; Pullan et al. 2008; de Silva et al. 2003), and was observed at the individual-level for hookworm in Chapter 5. However, there has been relatively little work to explore whether poorer households have any greater risk of infection per se. These models allow us to answer questions about infectious disease risk at a range of scales, and we used the index of SEP described in Chapter 5, and derived using the multiple factor analysis (MFA), as a predictor of household occupancy.

iii) Environmental covariates

Environmental conditions define the broad geographic distribution of areas of STH endemicity (Brooker et al. 2006b), but may also influence transmission dynamics at small spatial scales through heterogeneous effects on survival and level of environmental

contamination of free-living parasitic stages. The increasing availability of high resolution satellite imagery provides a means with which to explore these relationships (Pullan et al. 2012). The normalised difference vegetation index (NDVI), for example, a proxy measure of vegetation cover, is linked to soil moisture: all of the STH eggs rely on soil moisture for development, which occurs at faster rates with higher humidity (Brooker et al 2006b). Positive associations between NDVI and risk of individual infection with *Ascaris* (Saathoff et al. 2005) and hookworm (Pullan et al. 2008) have been reported from single-community based studies. Similarly, temperature is critical in the development of each of the STH species, and the distribution of each is associated with satellite derived measures of land surface temperature (LST) (Brooker et al. 2006b). Even within a species' temperature 'comfort zone', small-scale variation in land surface temperature could be expected to affect the rate of hatching and survival of hookworm larvae and embryonation of *Ascaris* and *Trichuris* ova, and may therefore impact upon parasite transmission dynamics (Anderson 1982). Variability in LST within small spatial areas was found to be associated with individual risk of hookworm (Riess et al. 2013) and *Ascaris* infection (Schüle et al. 2014) in Tanzania.

We used Fourier processed MODIS mean NDVI and maximum LST for the period 2001 to 2005 (data from Scharlemann et al. (2008)) as a predictor of household occupancy with each parasite. These data have a 1km spatial resolution, and the raster value at the location of each household was extracted using ArcMap 10.1.

iv) HIV infection in the household

The spatial heterogeneity in household HIV risk observed in chapter 4 was quite striking, and there was some overlap in high risk areas for *Trichuris* and, to a lesser extent, *Ascaris*. In order to explore whether the presence of HIV within a household could act as a predictor of occupancy for each of the STH infections, we included the presence of at least one individual with HIV infection as a binary predictor at the household-level (i.e. household-level infection with HIV, which we assume is observed perfectly).

v) Household size

A potentially important predictor of household infection is likely to be the number of opportunities for exposure outside the household, or the household population size.

vi) Latitude and Longitude

In Chapter 2, we identified broad spatial trends in STH risk, particularly for *Trichuris* and *Ascaris*, for which household-level infection appeared to be most common in the south and south eastern parts of the study area, respectively. We therefore included the x and y co-ordinates of study households as covariates in order to account for the observed large scale geographical gradients, as well as for landscape-based effects that were not captured by the other environmental and demographic variables.

Covariates on probability of detection

Each of the covariates described above (population density, SEP, NDVI, maximum LST, household HIV infection, household population size, latitude and longitude) was also included as a predictor of detection. This allowed us to explore whether these factors perform best in estimating the probability of occupancy (i.e. presence of absence of a parasite within a household) or detection (i.e. the probability of detecting infection in an individual given that they live in an occupied household), or both.

In the case of household-level HIV infection, we used the count of HIV positive individuals present to model detection rather than the binary "at least one" used for occupancy. HIV infection at the individual level was also included as a predictor of detection. Aggregated group level variables can be highly collinear with the individual-level predictors (that make up the aggregate) (Kreft & de Leeuw 1998) and we therefore paid particular attention to the stability of coefficients on individual- and household-level HIV infection when both were present as predictors of detection (model building strategies are described in the next section).

Age was included together with a quadratic term to account for the potentially non-linear relationship with each outcome. Gender and the presence of anaemia (on the basis of haemoglobin concentration: see chapter 3) were also included as possible predictors of detection at the individual level.

All continuous predictors of both occupancy and detection were scaled and centred (using the grand mean) to assist interpretation (Schielzeth 2010).

Table 6.1. Household and individual-level predictors for the probability of occupancy (ψ) and detection (*p*).

Household-level	**Process**	**Range of values**
Latitude	*p*, ψ	33.96 - 34.53 (degrees)
Longitude	*p*, ψ	0.059 – 0.77 (degrees)
Maximum land surface temperature	*p*, ψ	24 - 37°C
Mean NDVI	*p*, ψ	0.15 – 0.17
Household population size	*p*, ψ	1 to 30 people
Socioeconomic status	*p*, ψ	-2.97 to 4.59 (see Chapter 5)
HIV infection in household	ψ	0/1
HIV count in household	*p*	0 – 4 infected people
Local population density	*p*, ψ	22 – 2,811 per km^2
Local population density*Local population density	*p*, ψ	484 – 7,901,721 per km^2
Individual-level		
HIV infection	*p*	0/1
Age	*p*	5 to 85 years
Age*Age	*p*	25 – 7225 years
Male gender	*p*	0/1
Anaemia	*p*	0/1

6.2.3. Model selection

Zero inflated binomial models were fit using the *occu* function in *unmarked* (Fiske & Chandler 2011) in R.

For each parasite, we fit and present the full model containing the combination of all household-level covariates as predictors of both detection and occupancy, and all individual-level covariates as predictors of detection. The importance of including quadratic terms on age and population density in the full model was assessed on the basis of information criteria, with inclusion of the quadratic term in each case if it resulted in lower AIC. We used the purposeful selection procedure (Hosmer et al. 2013) to select the best model in terms of fit and parsimony. This approach involves the step-wise removal of predictors from the full model on the basis of the alpha-level of the Wald test, together with careful examination of the effect of variable removal on the remaining coefficients that might suggest a confounding effect. Improvements in model fit were assessed using a likelihood ratio test at each step. The final model selected on the basis of purposeful selection contained those variables significant at traditional alpha levels (i.e. 5%), and/or those that appear to have an important confounding effect.

We followed Kery et al. (2010), and initially selected the best model in terms of the detection process (i.e. in occupied households) followed by the best model in terms of occupancy (containing the best fitting predictors of detection). Having selected the best

overall model, all predictors not already present were re-introduced to both the detection and occupancy parts of these two part models, with assessment of the effect of variable inclusion on coefficients and Wald p-values.

Goodness-of-fit of the full and best fitting models for each parasite were assessed using the parametric bootstrap approach described by MacKenzie & Bailey (2004).

6.2.4. Estimating the proportion of sites occupied

In chapter 4, we adopted several approaches to derive an estimate of household-level prevalence, or the proportion of sampled households that contained at least one infected individual. The zero inflated binomial models described here are somewhat analogous to Bayesian binomial mixture models described in that chapter (with the added advantage that the probabilities in the mixture models are linked to regression equations), and can also be used to derive an estimate of household prevalence of infection. The probability of occupancy, ψ, following adjustment for any covariates, for example, will give an estimate of the probability that a randomly selected household in the population under study is occupied by the species of interest. Moreover, for those households we sampled but in which infection was undetected, we can use Bayes Theorem together with household specific estimates of ψ (i.e. adjusted by covariates), the total number of people (*j*) sampled (*T*) together with the probability of detection (*p*) for each person (again, conditioned on covariates) in order to estimate the probability that the *i*th household is truly infected as (MacKenzie et al. 2006):

$$\Pr(z_i = 1 | \, y_i = 0) = \frac{\psi \prod_{j=1}^{T}(1-p)}{(1-\psi) - \psi \prod_{j=1}^{T}(1-p)}$$

We summed over these conditional probabilities of household infection (if the parasite was unobserved: the probability is always 1 if it is observed) to give an estimate of the proportion of sites that can be considered to be infected, and used parametric bootstrap methods to estimate confidence intervals. The resulting summary is akin to the household-level prevalence (i.e. proportion of households sampled that are infected), with adjustment for factors that influence the probability of occupancy and the probability of detection given occupancy.

6.3. Results

Here we present the best fitting model selected on the basis of purposeful selection, as well as the results from the full model for comparison. The steps in the model selection procedure, including univariate results, are given in Appendix 5.1.

6.3.1. Hookworm

The goodness of fit test yielded a p-value of 0.153 after 1000 bootstrap replicates for the full model describing household hookworm infection, and 0.31 for the model selected by purposeful selection, indicating no evidence of a lack of fit.

Probability of occupancy

The adjusted estimates of household-prevalence based on the full site occupancy model and that selected by purposeful selection represent a moderate increase from the raw value of 69.6% (95% CI 64.8 – 73.9) to around 80% (Table 6.2). These estimates are broadly equivalent to that derived using the hyper-geometric approach described in Chapter 4 (78%). Households in which infection was unobserved, but which were classified as infected on the basis of the occupancy model (i.e. where $\Pr(z_i = 1 | y_i = 0) > 0.5$), are shown in Figure 6.1. We used the kernel smoothing approach (described in chapter 4) to present the resulting change in the spatial distribution of household-level risk of infection (briefly, we derived a smoothed surface for all positive cases, and divided that by a smoothed surface for the population at risk. Both surfaces used a bandwidth of 5km with correction for edge effects and were implemented in the *sparr* package in R).

The NDVI was a positive predictor of the likelihood that a household is occupied by hookworm. Holding all other predictors of occupancy at their average values, the probability that a household in an area with the highest NDVI (i.e. most densely vegetated) in our study area is occupied ("infected") by hookworm would be predicted to be 0.93 (0.84 – 0.97), whilst the probability of occupancy in the least vegetated area would be 0.41 (95% CI 0.17 – 0.70).

There was also some weak evidence for a negative effect of increasing population density on the likelihood that a household is occupied by hookworm (the likelihood ratio test gave a p-value of 0.06 following the removal of this term). This variable was retained together with household count which had a small (confounding) effect on the co-efficient for NDVI (reducing it from 0.54 to 0.48).

Probability of detection

For those people living in a household that is occupied by hookworm, age, sex and HIV infection status appeared to be important factors operating at the individual level on the probability of detection (a mixture of the probability that that individual is infected given that they live in an occupied household, and the probability we detect the infection given that they are infected) (Table 6.2).

Holding all predictors in the 'best' model (identified through purposeful selection) at zero, the average probability of detection of hookworm on the basis of microscopic examination of a single faecal sample from a female of average age (25 years) living in a hookworm occupied household would be 0.45 (95% CI 0.40 - 0.50). If she was infected with HIV, the probability of detection is reduced to 0.19 (95% CI 0.12 – 0.29). The probability of detection in a 25 year old male, by contrast, would be 0.64 (95% CI 0.58 – 0.70), and 0.24 (95% CI 0.15 – 0.36) if he was HIV positive.

The best fitting model included a negative quadratic term for age, hence we would expect a lower probability of detection in a female (0.42 (95% CI 0.24 – 0.62) or male (0.49 (95% CI 0.30 – 0.69)) of 85 years without HIV, as well as in a girl (0.36 (95% CI 0.31 – 0.41) or boy (0.43 (95% CI 0.38 – 0.49)) of 5 years of age.

Socioeconomic position appeared to influence the probability of detection in an occupied household rather than the presence or absence of infection. Holding all other predictors at zero, we would expect the probability of detection in a person sampled within the poorest household in our study area to be 0.66 (95% CI 0.58 – 0.73), whilst it would be 0.18 (95% CI 0.12 – 0.26) in the richest household (given that it is occupied by the parasite). The number of people living within a household was also a positive predictor of the probability of detection, as was latitude.

We can use these estimates to derive an indication of the sample size required in order to achieve some level of confidence of detecting at least one individual with hookworm, given that a household is occupied by the parasite. The average probability of detection (i.e. holding all predictors at zero) in the model selected by purposeful selection was 0.45 (95% CI 0.40 – 0.50). Hence, it would be necessary to sample 5 or more people to ensure 95% confidence of detecting at least one person with hookworm (i.e. 1-(1-0.45)^5 = 0.95). This scenario is presented in Figure 6.2, together with the effect of varying the value of household-level predictors of detection.

Table 6.2. Output from the full site occupancy models and that selected by purposeful selection (PS) for hookworm for the effect of covariates on probability of household infection ψ and probability of detection given household infection, p

	p (Full) Estimate (95% CI)	ψ (Full) Estimate (95% CI)	p (PS) Estimate (95% CI)	ψ (PS) Estimate (95% CI)
Intercept	-0.16 (-0.38, 0.049)	1.46 (1.06, 1.85)	-0.19 (-0.39, 0.011)	1.53 (1.15, 1.91)
Individual-level				
Age	0.61 (0.45, 0.76)	-	0.60 (0.45, 0.76)	-
Age^2	-0.20 (-0.31, -0.090)	-	-0.20 (-0.30, -0.09)	-
Male	0.32 (0.10, 0.53)	-	0.32 (0.11, 0.53)	-
HIV	-1.33 (-1.94, -0.73)	-	-1.28 (-1.84, -0.72)	-
Anaemia	-0.019 (-0.27, 0.23)	-	-	-
Household-level				
SEP	-0.43 (-0.57, -0.29)	-0.076 (-0.49, 0.34)	-0.42 (-0.54, -0.29)	-
NDVI	0.080 (-0.06, 0.22)	0.44 (0.082, 0.79)	-	0.48 (0.16, 0.81)
LST	-0.01 (-0.17, 0.16)	-0.021 (-0.42, 0.38)	-	-
HIV	0.020 (-0.11, 0.16)	0.14 (-0.83, 1.10)	-	-
Pop. dens.	0.024 (-0.17, 0.22)	-0.32 (-0.69, 0.047)	-	-0.33 (-0.65, 0)
Household count	0.12 (0.001, 0.24)	0.29 (-0.13, 0.70)	0.14 (0.033, 0.24)	0.25 (-0.11, 0.60)
Longitude	-0.059 (-0.20, 0.084)	-0.096 (-0.51, 0.32)	-	-
Latitude	0.28 (0.11, 0.45)	0.26 (-0.17, 0.69)	0.25 (0.13, 0.38)	-
Household prevalence	79.8% (95% CI 77.4, 82.1)		80.3% (95% CI 78.1, 82.8)	
AIC	2340		2323	

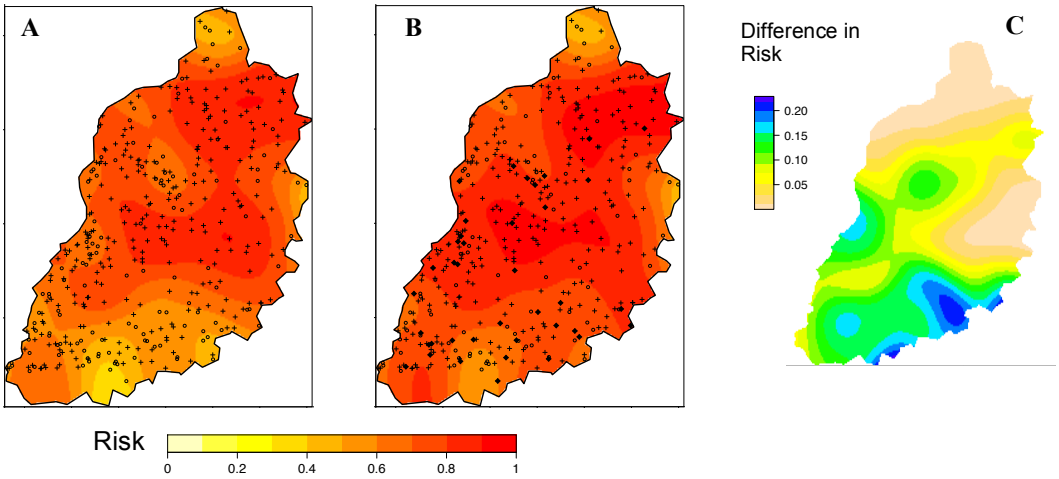


Figure 6.1. Spatial distribution of household-level risk of hookworm infection: raw (unadjusted) estimates (A); adjusted based on best fitting occupancy model (B) (+ are positive households, o are negative households, ◆ are households in which the status changed from negative to positive). Estimates in C show the difference in risk estimate between adjusted and raw estimates of risk.

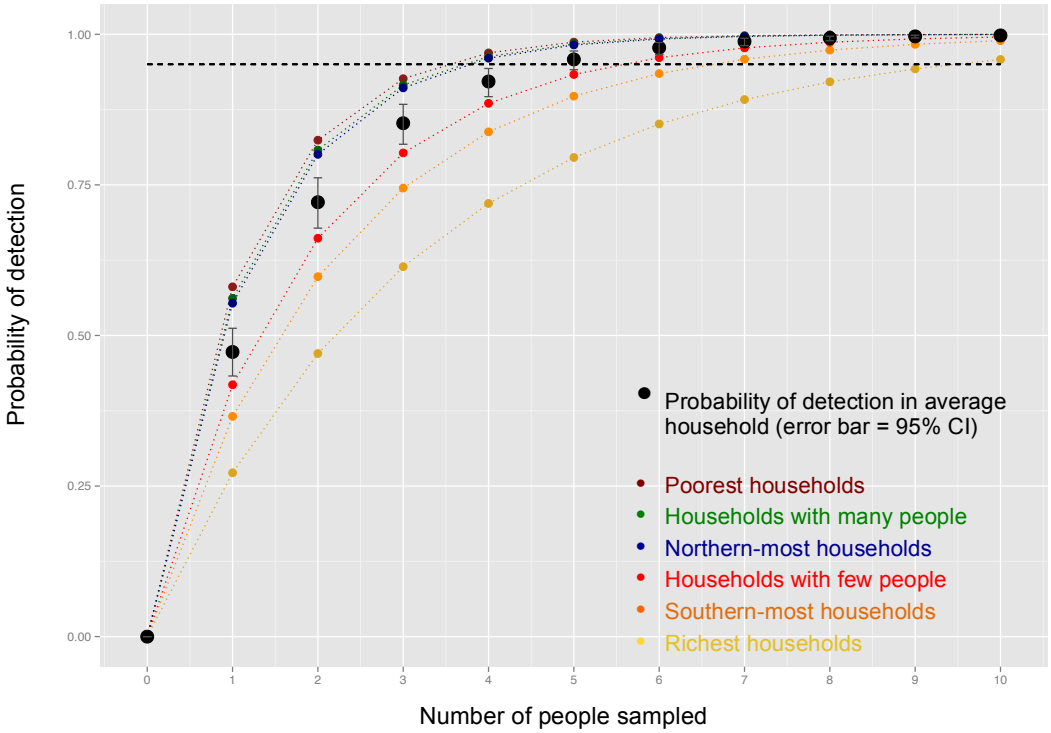


Figure 6.2. Probability of detecting at least one individual with hookworm in an 'average' household in which hookworm is present on the basis of variable sampling effort and in which all predictors from the best fitting model (Table 6.2) are held at zero. Coloured points represent probability of detection when the listed variable is assigned the value at its 5th or 95th percentile and all other variables are held at zero.

6.3.2. *Ascaris lumbricoides*

The goodness of fit test yielded a p-value of 0.42 after 1000 bootstrap replicates for the full model, and 0.34 for the model selected by purposeful selection, indicating no evidence of a lack of fit.

Probability of occupancy

The full and best fitting model based on purposeful selection resulted in a household level prevalence of around 40% (Table 6.3). This is considerably higher than the raw estimate of 23.8% (95% CI 19.9 – 28.3), but is similar to that derived using the hyper-geometric approach described in chapter 4 (41%). Latitude and longitude were negative and positive predictors of occupancy, respectively, hence the majority of households whose status changed from negative to positive are in the south-east of the study area (Figure 6.3).

Probability of detection

Only age was a significant predictor of detection, with the effect decreasing linearly (Table 6.3), so that we would expect the probability of detection of *Ascaris* in a child (of either sex, since gender did not appear in the model selected by purposeful selection) of 5-10 years living in a household occupied by the parasite to be 0.31 (95% CI 0.25 – 0.38), whilst the probability in a 85 year old in the same household would be just 0.048 (0.022 – 0.10).

The NDVI was also a positive predictor of the probability of detection (having no effect on occupancy), whilst household SEP was a negative predictor. We would expect the probability of detection to be 0.39 (95% CI 0.27-0.52) in the poorest household in our study area, and just 0.071 (95% CI 0.035-0.14) in the richest.

Interestingly, whilst there was no evidence of an effect of individual HIV infection on the probability of detection (although the co-efficient was reasonably strongly negative), the probability of detection of *Ascaris* within an occupied household appeared to increase as the count of HIV positive individuals within it increases. This effect persisted when HIV was not included in the model selected by purposeful selection ($\beta = 0.16$ (95% CI 0.01 to 0.32)), but the inclusion of HIV infection status at the individual level reduced the value of the co-efficient of the aggregated variable by more than 20%, so we retained it. It may be useful to note that HIV count has no effect on the probability of detection at the univariate-level (see Appendix 5.1), and it appears to be the inclusion of control for SEP that results in the observed effect.

The average probability of detection (i.e. holding all predictors at zero) was 0.22 (95% CI 0.18 – 0.26). The sample size necessary to identify *Ascaris* in an infected household on the basis of single faecal samples would therefore be around 13 people. The variable effect of household-level predictors on this estimate is shown in Figure 6.4, and suggests that whilst in the poorest households one would need to sample 8 people to have a reasonable degree of confidence of identifying at least one with *Ascaris* infection, this would need to be more than 20 in the richest households.

Table 6.3: Output from the full site occupancy models and that selected by purposeful selection (PS) for *Ascaris lumbricoides* for the effect of covariates on probability of household infection ψ and probability of detection given household infection, p

	p (Full model) Estimate (95% CI)	ψ (Full model) Estimate (95% CI)	p (PS) Estimate (95% CI)	ψ (PS) Estimate (95% CI)
Intercept	-1.22 (-1.62, -0.82)	-0.42 (-0.98, 0.14)	-1.29 (-1.54, -1.04)	-0.53 (-0.88, -0.17)
Individual level				
Age	-0.51 (-0.74, -0.28)	-	-0.52 (-0.74, -0.30)	-
Male	-0.28 (-0.65, 0.08)	-	-	-
HIV	-0.64 (-1.57, 0.29)	-	-0.61 (-1.53, 0.31)	-
Anaemia	-0.041 (-0.45, 0.37)	-	-	-
Household level				
SEP	-0.47 (-0.76, -0.19)	0.20 (-0.24, 0.70)	-0.40 (-0.62, -0.18)	-
NDVI	0.21 (-0.072, 0.47)	0.019 (-0.40, 0.44)	0.22 (0.012, 0.43)	-
LST	-0.044 (-0.26, 0.17)	-0.21 (-0.58, 0.16)	-	-
HIV	0.24 (0.03, 0.45)	0.022 (-0.94, 0.99)	0.20 (0.032, 0.37)	-
Pop. dens	-0.072 (-0.37, 0.22)	0.19 (-0.31, 0.69)	-	-
Household count	0.12 (-0.18, 0.42)	-0.16 (-0.60, 0.28)	-	-
Longitude	0.016 (-0.29, 0.32)	1.05 (0.47, 1.64)	-	0.84 (0.42, 1.27)
Latitude	0.11 (-0.22, 0.45)	-1.25 (-1.89, -0.61)	-	-1.18 (-1.66, -0.69)
Household prevalence	41.8% (38.6, 44.7)		39.3% (35.9, 42.3)	
AIC	1026		1008	

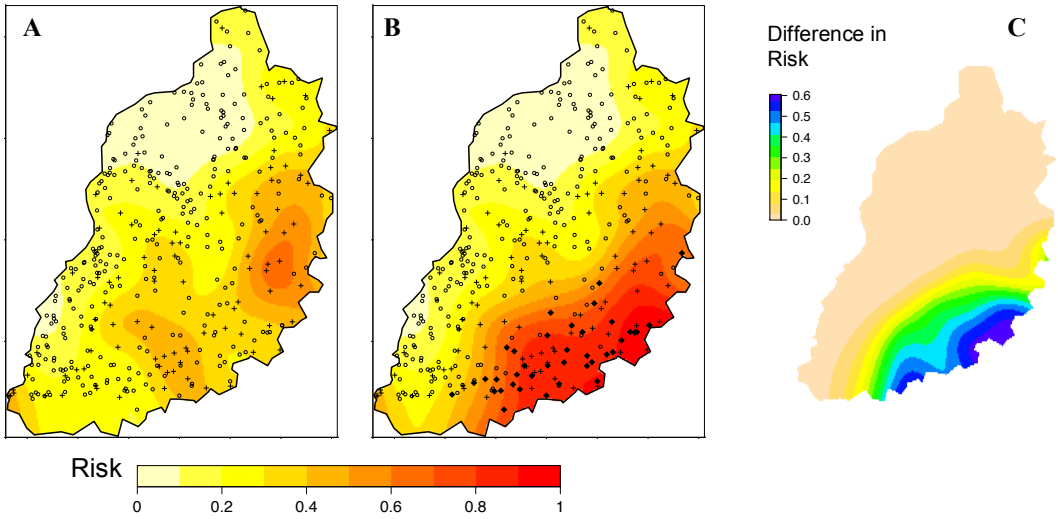


Figure 6.3. Spatial distribution of household-level risk of *Ascaris* infection: raw (unadjusted) estimates (A); adjusted based on best fitting occupancy model (B) (+ are positive households, o are negative households, ◆ are households in which the status changed from negative to positive). Estimates in C show the difference in risk estimate between adjusted and raw estimates of risk.

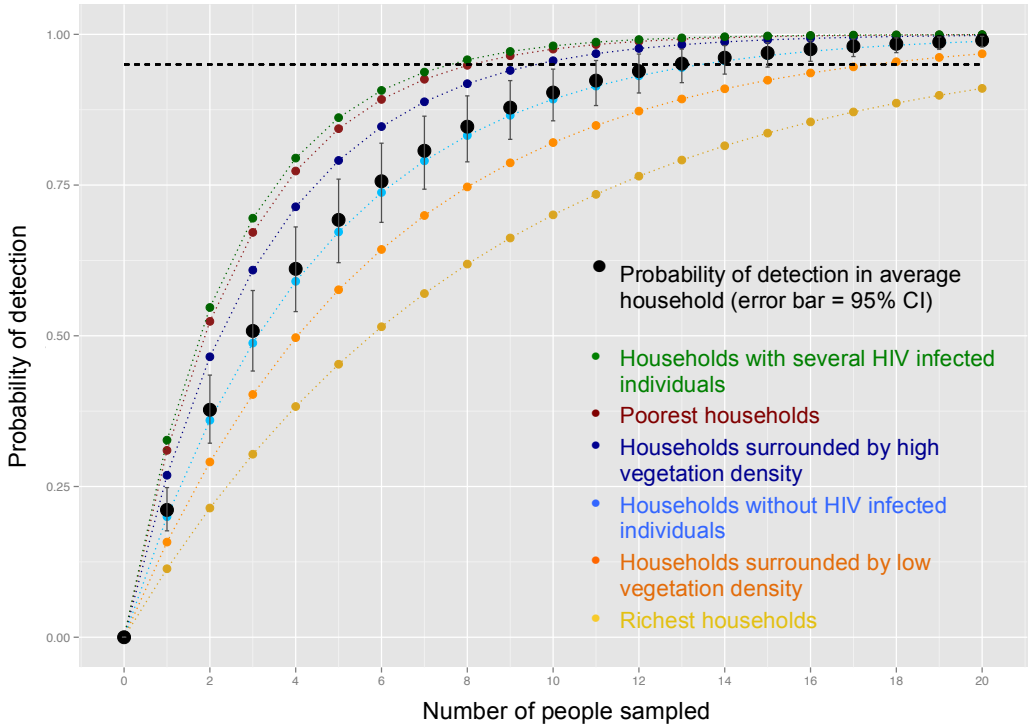


Figure 6.4. Probability of detecting at least one individual with *Ascaris* in an 'average' household in which the parasite is present, with variable sampling effort, and in which all predictors from the best fitting model (Table 6.3) are held at zero. Coloured points represent probability of detection when the listed the variable is assigned the value at its 5th or 95th percentile, and all other variables are held at zero.

6.3.3. *Trichuris trichiura*

The goodness of fit test yielded a p-value of 0.57 after 1000 bootstrap replicates for the full model, and 0.40 for the model selected by purposeful selection, indicating no evidence of a lack of fit.

Probability of occupancy

The raw (unadjusted) estimate of the proportion of households occupied by *Trichuris* was 23.8% (95% CI 19.9-28.3). This increases quite substantially to more than 50% in both the full site occupancy model and that selected by purposeful selection (Table 6.4) (the household prevalence estimate in chapter 4 was 46%). This effect is influenced in large part by the very strong relationship between latitude and probability of occupancy: virtually all of the households in the southern half of the study area in which *Trichuris* was not detected are predicted to be occupied by the parasite (Figure 6.5). Increasing maximum land surface temperature also had a positive effect on the probability that a household was occupied.

Probability of detection

At the individual-level, age, sex and HIV status were all negative predictors of detection probability for *Trichuris* (Table 6.4). Holding all other predictors constant, the probability of detection in a female of 5-10 years of age without HIV would be 0.30 (95% CI 0.24 – 0.37), whilst it would be just 0.027 (95% CI 0.011- 0.066) in a 85 year old male with HIV. The individual effect of HIV was quite substantial, and the probability of detection in a 25 year old female infected with the virus is expected to be 0.09 (95% CI 0.04 to 0.18), considerably lower than if she was HIV negative (0.25 (95% CI 0.21 – 0.30))

With a similar effect to that observed for *Ascaris*, count of HIV positive individuals in a household increased the probability of detection of *Trichuris*. The best fitting models included HIV infection at the individual level and the aggregated count of HIV infected individuals at the household-level. There was some evidence of a negative confounding effect for both variables. For individual HIV infection, the coefficient in the model selected by purposeful selection, but without the aggregated HIV variable, was more than 20% of that in the model containing both variables (-0.91 (95% CI -1.64, -0.19) vs. -1.22 (95% CI -1.98, -0.46)). In the absence of account for individual HIV infection, the aggregated variable became non-significant and its co-efficient was reduced by more than 50% (to 0.085 (95% CI -0.05-0.22)). This negative confounding could perhaps be expected (Mehio-Sibai et al.

2005): individual HIV infection is negatively associated with the outcome, household HIV count is positively associated (and both effects are significant at the univariate level, see Appendix 5.1), but they are (inherently) positively related with each other.

While LST was a positive predictor of occupancy, it seems to have the reverse effect on the probability of detection. Somewhat counter-intuitively, the number of people resident in a household also has a negative effect on detection. This effect was not observed at the univariate level (Table 6.4). Local population density appears to increase the probability of detection in an occupied household (but not household infection), with model fit being improved with the inclusion of a quadratic term.

Holding all predictors in the model selected by purposeful selection at zero, the average probability of detection of *Trichuris* in a household occupied by the parasite was 0.252 (95% CI 0.21 – 0.30). In order to detect at least one person with *Trichuris* on the basis of a single faecal sample with 95% confidence, it would therefore be necessary to sample at least 11 people. The effect of varying the continuous household-level predictors of detection on this estimate are shown in Figure 6.6.

Table 6.4. Output from the full site occupancy models and that selected by purposeful selection (PS) for *Trichuris trichiura* for the effect of covariates on probability of household infection ψ and probability of detection given household infection, p

	p (Full) Estimate (95% CI)	ψ (Full) Estimate (95% CI)	p (PS) Estimate (95% CI)	ψ (PS) Estimate (95% CI)
Intercept	-1.00 (-1.43, -0.57)	0.13 (-0.81, 1.08)	-1.09 (-1.35, -0.83)	0.28 (-0.39, 0.96)
Individual level				
Age	-0.23 (-0.40, -0.068)	-	-0.23 (-0.40, -0.07)	-
Male	-0.52 (-0.86, -0.19)	-	-0.51 (-0.83, -0.18)	-
HIV	-1.26 (-2.02, -0.49)	-	-1.22 (-1.98, -0.46)	-
Anaemia	0.075 (-0.28, 0.43)	-	-	-
Household level				
SEP	0.16 (-0.088, 0.40)	-0.47 (-1.08, 0.15)	-	-
NDVI	0.12 (-0.067, 0.30)	-0.27 (-0.94, 0.40)	-	-
LST	-0.32 (-0.51, -0.13)	0.91 (0.20, 1.62)	-0.25 (-0.42, -0.08)	0.90 (0.39, 1.41)
HIV	0.19 (0.033, 0.35)	0.32 (-0.90, 1.54)	0.18 (0.037, 0.33)	-
Pop. dens	0.52 (0.20, 0.85)	-0.058 (-0.66, 0.54)	0.50 (0.21, 0.79)	-
Pop. dens^2	-0.075 (-0.15, 0.01)	-	-0.074 (-0.15, -0.001)	-
Household count	-0.29 (-0.48, -0.11)	0.54 (-0.14, 1.22)	-0.17 (-0.33, -0.022)	-
Longitude	0.069 (-0.16, 0.30)	0.21 (-0.47, 0.90)	-	-
Latitude	0.086 (-0.30, 0.47)	-3.33 (-4.76, -1.90)	-	-3.14 (-4.36, -1.92)
Household prevalence	51.1% (48.6 - 53.3)		52.8% (50.1 - 55.0)	

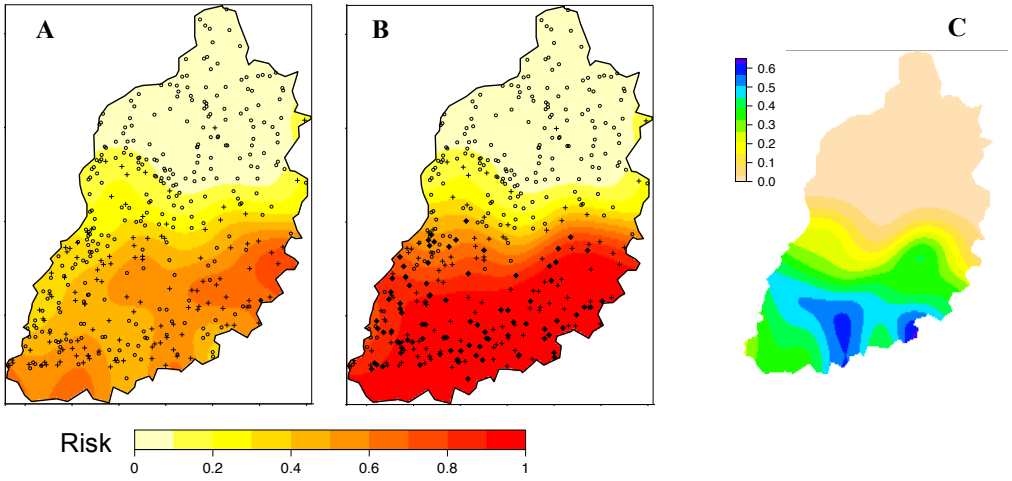


Figure 6.5. Spatial distribution of household-level risk of *Trichuris* infection: raw (unadjusted) estimates (A); adjusted based on best fitting occupancy model (B) (+ are positive households, o are negative households, ◆ are households in which the status changed from negative to positive). Estimates in C show the difference in risk estimate between adjusted and raw estimates of risk.

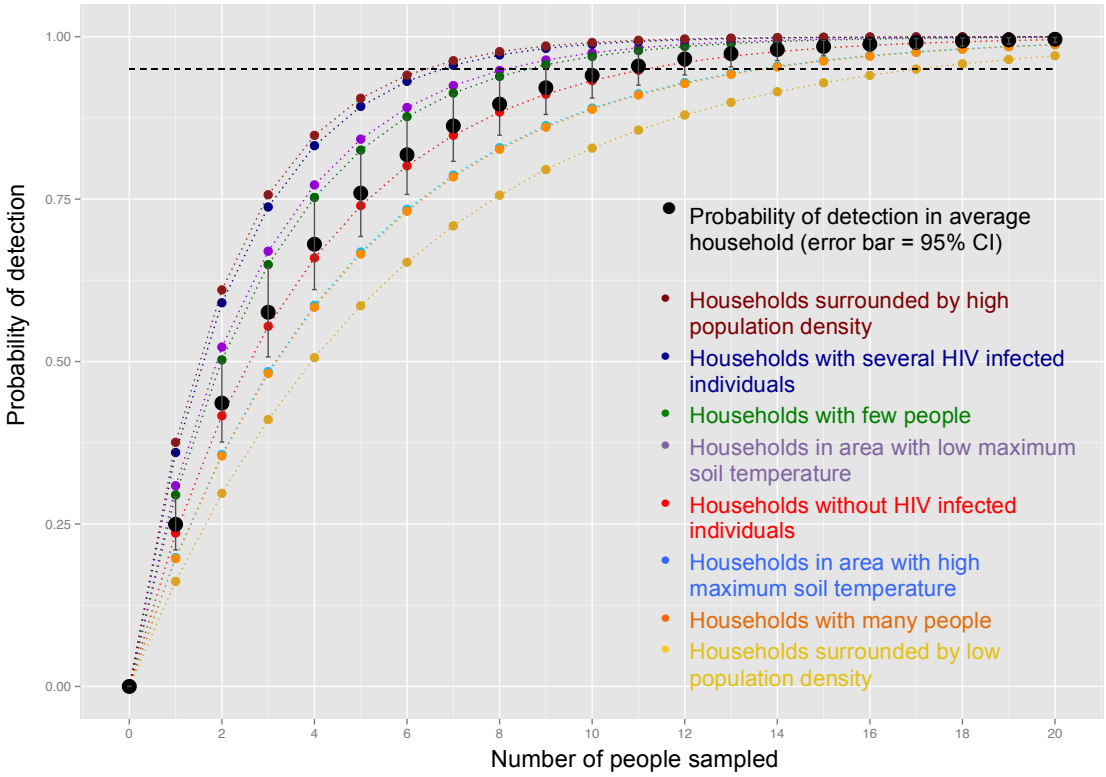


Figure 6.6. Probability of detecting at least one individual with *Trichuris* in an 'average' household in which the parasite is present, with variable sampling effort, and in which all predictors from the best fitting model (see Table 6.4) are held at zero. Coloured points represent probability of detection when the listed the variable is assigned the value at its 5th or 95th percentile, and all other variables are held at zero.

6.4. Discussion

In this chapter, we have considered the dichotomy between the presence and absence of STH infection at the household-level to be an outcome of substantive interest. By exploring predictors of infection at this level, with control for the observation process in sampled individuals, we have been able to derive estimates of between-household prevalence that are adjusted on the basis of the likelihood that a household is 'infected' by a parasite given that infection was unobserved in sampled individuals. The results suggest that infection prevalence (at least at the household level) is underestimated, particularly for *Trichuris* and *Ascaris*, where the adjusted estimates were more than double the observed (and, encouragingly, not too far off the adjusted estimates for these parasites derived in chapter 4). That the prevalence of helminth infections may be biased using highly imperfect tests and on the basis of a single faecal sample is not new (Booth et al. 2003; Tarafder et al. 2010; de Vlas & Gryseels 1992), but, to our knowledge, this is the first application of zero-inflated binomial regression to derive prevalence estimates that are corrected for detection.

Whilst the approach described here is reasonably straightforward, and could be implemented in any number of statistical packages and dedicated ecological software (e.g. PRESCENCE[7] or MARK[8]) (or even Microsoft Excel using the solver function), there are some important sources of error that need to be considered. Potentially most important is that these models (and the underlying binomial sampling procedure) assume that each survey (person sampled per household) is independent (MacKenzie et al. 2006). As we described in chapter 4, infection with the STH is highly clustered at the household level, and therefore this independence assumption is not met. The issue of heterogeneous detection probabilities between sites is widely recognised in the site occupancy modelling (and ecological) literature (MacKenzie et al. 2006; Royle 2006; Royle & Nichols 2003), where it is generally attributed to varying abundance of the species of interest (Tanadini & Schmidt 2011). Several authors have suggested that including site level predictors of species abundance may be sufficient to account for the expected heterogeneities in detection per site (Kery et al. 2013; Schmidt et al. 2013). We have sought to do similar here by including effects that we expect will influence between-household variation in individual infection risk, and therefore the probability of detection in occupied households. However, it seems unlikely that all of this variation can be accounted for, and a particularly important source of unmodelled heterogeneity in detection probability (i.e. individual infection within occupied households) may be the genetic relatedness of household members. Genetic dispositions are likely to

[7] http://www.mbr-pwrc.usgs.gov/software/presence.html
[8] http://warnercnr.colostate.edu/~gwhite/mark/mark.htm

impact on helminth susceptibility (Quinnell 2003; Chan et al. 1994a), and several studies have revealed that shared genetic effects are important in explaining between household variability in infection intensity (Bethony et al. 2002; Williams-Blangero et al. 1999).

Residual dependency in observations can lead to well-known issues in the underestimation of parameter variance, so that null hypotheses are too easily rejected (McDermott & Schukken 1994), or in estimates that may be biased (i.e. too large or too small) (MacKenzie et al. 2006; Royle 2006). The extent to which these sources of error may impact inference in our own models is uncertain, however the goodness of fit test used here (and which indicated no evidence for lack of fit) is reported to have some power to detect lack of fit in site occupancy models that is caused by detection histories that are not independent (MacKenzie & Bailey 2004).

A sensible and reasonably straightforward solution would be to include a random effect at the household level in order to account for any un-modelled heterogeneity in detection probability (Royle 2006). The inclusion of random effects is not currently available in the packages for running site occupancy models within a maximum likelihood framework, but is reasonably straightforward in a MCMC-based implementation within WinBUGS or similar. Work is underway to examine the effect on inference of including a random intercept (with either a beta-binomial or logit-normal distribution (Royle 2006)) for each household.

A second potential source of error may be residual (i.e. unaccounted for) spatial dependency in household infection (occupancy) status. Exploring residuals for spatial autocorrelation is not straightforward in these models, since the outcome being modelled (household infection) is, in many cases, unobserved. As is the case for within household correlation, we do not know the extent to which residual spatial autocorrelation (which may affect estimated coefficients and errors (Kuhn 2007; Kuhn & Dormann 2012)) may impact upon inference. However, we included a range of household level predictors that we expected would explain spatial dependency in household level infection, including latitude and longitude to account for broad geographic trends (and which had a major effect on model fit (in terms of AIC) for both *Ascaris* and *Trichuris* (see Appendix 5.1)). Performing our analysis within a Bayesian framework will potentially allow us to account for spatial autocorrelation in a more rigorous way, for example through the use of autologistic models (Royle & Dorazio 2008; Burton et al. 2012; Bled et al. 2013).

In our application of site occupancy models, "detection" is a mixture of two probabilities: the probability that an individual is infected given that they live in an occupied household, and the probability that we detect infection given that they are infected. These probabilities

are difficult to separate since we only collected a single faecal sample from each person. However, where multiple samples are available from the same individual, or multiple (independent) tests can be performed on the same sample, it would be possible to consider occupancy (infection) and detection as separate processes (probabilities) at the individual level. In such a case, the probability of detection could be considered analogous to diagnostic sensitivity (per sample in the case of repeat samples or per test in the case of repeat tests) (Lachish et al. 2011). This application of ZIB models would allow individual prevalence estimates to be adjusted on the basis of a sensitivity (probability of detection) estimate that is derived from within the population of interest, and which could be conditioned by covariates, and in a very straightforward way. This is a potentially powerful application of these models. Furthermore, such a model could be extended (within WinBUGS or similar) to three-levels (Schmidt et al. 2013; Mordecai et al. 2011; Nichols et al. 2008), and could therefore consider occupancy (infection) as substantive outcome at both the individual and household level.

It would also be possible to extend the two-level hierarchical models described in this chapter to three-levels by considering the infection intensity at the individual level (on the basis of eggs per gram of faeces), rather than the presence or absence of (detectable) infection. Chipeta et al. (2013) used a zero inflated negative binomial approach to explore factors that predict individual infection with *Schistosoma haematobium* in Zambian school children, and the factors that influence intensity of infection given infection. It would be possible to use a similar model to consider the factors that influence intensity of infection given that an individual is infected, and given that they live in an infected household (with zero inflation at the individual and household levels). This would be a particularly efficient use of our data, that would allow exploration of a range of outcomes (individual infection intensity, individual infection and household infection).

These models have allowed us to ask a number of interesting questions about the factors that may influence whether or not a household is “infected” or occupied by a parasite, and those factors that influence the likelihood that we detect infection in individuals living in occupied households.

We repeat the findings in chapter 5 by showing potentially important relationships between HIV and the STH. For each species, HIV infection at the individual level was associated with a reduced probability of detection, although this effect was only significant for hookworm and *Trichuris*. These effects were quite substantial, such that the probability of detecting hookworm or *Trichuris* in a HIV infected individual living in a household in which

the parasite is present was less than 50% of that if the person did not have HIV (whilst controlling for age and sex). This suggests HIV infected individuals may be less likely to be infected, or are less likely to shed eggs in sufficient quantities to allow detection (due to an infection intensity or a fecundity related process).

In the introductory chapter we described how immune dysregulation associated with helminth infection may increase individual susceptibility to HIV. We might therefore have expected that HIV infected individuals would have a higher prevalence of helminth infection. It is interesting to note, however, that the negative relationship observed between HIV and STH infection in this study has also previously been reported by a number of other studies (Brown et al. 2006). To explore this further, we performed a small meta-analysis of those studies that compare helminth prevalence in HIV positive individuals with HIV negative controls. The methodology and results are given in Appendix 6, and whilst the quality of the available studies and their settings were highly variable, we found that HIV infected individuals tended to have a lower prevalence of infection for each of the STH than those without HIV. This effect was only statistically significant in the case of hookworm, although there was also weak evidence to support the effect in the case of *Ascaris*. The explanation for this trend, and the strong effect seen in our own data, is uncertain, but it might be the case that the persistence or establishment of gut parasites (or their larval migrans (Viney et al. 2004)) is reduced in HIV infected individuals, who often have enteropathies (Morales et al. 1995; Hunter et al. 1992). Alternatively, such individuals may be just as likely to be infected by helminths, but HIV induced structural or immunological changes to the gut mucosa impact upon egg production and therefore detection (Karanja et al. 1997).

Despite the negative effect of HIV infection on the probability of detection at the individual level, we observed a positive relationship with the count of HIV positive individuals in the household for *Ascaris* and *Trichuris*. It would therefore appear that in those households in which the probability of detection for these parasites is high, the within-household transmission, or rate of external acquisition, of HIV is also highest. More work is needed to control for potential confounders, and particularly for unmodelled heterogeneity in between-household detection probability (as described above). However, a possible (and at this stage very tentative) hypothesis could be proposed that the high levels of within household-transmission of these parasites in occupied households (which would be expected to increase the probability of detection) might impact upon within household HIV transmission dynamics by increasing individual susceptibility to HIV, or to the infectiousness of HIV co-

infected individuals (Quinn et al. 2000; Brown et al. 2006). This is certainly an area for further work.

An additional interesting finding is that household poverty does not appear to impact on the probability that it is infected (or occupied) by any of the parasites of interest. Instead, low SEP appears to impact on the probability of detection (for hookworm and *Ascaris*) within infected households. This suggests that rich and poor households do not have a different risk of infection with these two parasites, but that there may be *more* transmission within poor households, so that more people are infected, or people have more intense infections and therefore the probability of detection is higher. To reach this conclusion, we have of course assumed that most transmission occurs within the domestic domain (Cairncross et al. 1996), but it may also be the case that poorer individuals also have greater exposure outside the home.

While the density of vegetation (NDVI) surrounding a household appears to influence whether or not it is occupied by hookworm, it performs better as a predictor of detection in the case of *Ascaris*, and has no influence on either occupancy or the observation process for *Trichuris*. Thus, households in areas with low NDVI are less likely to have hookworm infection than households in areas with high NDVI, whilst households in low NDVI areas are just as likely to be occupied by *Ascaris* as those households in high NDVI areas (or there is no evidence for a difference), but there is a tendency for fewer people to be infected, or for infections to be less intense (or for some other factor that reduces shedding of eggs), and therefore for a lower probability of detection. This difference in the effect of NDVI, influencing the dichotomy between none and some in the case of hookworm, and (presumably) the transmission intensity in occupied households in the case of *Ascaris*, is difficult to explain. However, it suggests that NDVI acts on transmission of hookworm at a broader spatial scale than for *Ascaris*, and may therefore explain hetereogeneity in community prevalence but may be more important for small-scale hetereogeneities in individual risk of infection in the case of *Ascaris*.

Land surface temperature was only important as a predictor of *Trichuris*, and operated at the level of both the household (occupancy) and individual (detection). These effects were opposite, so that there was a higher probability of occupancy in the warmer areas, but a lower probability of detection. Again, this effect is difficult to explain, but may suggest higher levels of between-household transmission in warmer areas (or a higher community prevalence) (and this might be confounded by some unmeasured factor), but that higher land surface temperatures does not favour transmission within the occupied household. The

optimum temperature for survival and development of *Trichuris* (and indeed for *Ascaris*, although no effect was observed in our study) is between 28 and 32°C, with development arresting beyond 38°C (Brooker et al 2006b). Average maximum temperatures in our study area range between 24 and 37°C, and hence it is reasonable to expect that parasite development, and the rates of within household transmission, may be reduced in the warmest parts. The ability to identify these sort of heterogeneous effects on the state (household infection) and detection (individual infection) process is part of the power of these ZIB models (Kery 2008).

Main findings

1. Household infection with each STH species, and particularly hookworm, is common in this mixed farming community in western Kenya, but the prevalence for all three species is likely to be underestimated based on the observed data;
2. SEP does not appear to impact upon the likelihood that a household is infected by each parasite, but greater levels of transmission seem to be more likely in poorer households;
3. There is a negative relationship between HIV infection and the probability of detection of hookworm and Trichuris in individuals in infected households, but in those households in which the probability of detection of *Ascaris* and *Trichuris* is high (and, we assume, within household transmission is therefore greatest) there is a tendency for more people to be infected by HIV;
4. Zero-inflated binomial models provide a straightforward means to explore infection status whilst controlling for the detection process.

Chapter 7

Shared household-level risk factors for neglected tropical diseases (NTDs) in a co-endemic population

7.1. Introduction

Polyparasitism is widely recognised to be the norm rather than the exception throughout the developing world (Buck et al. 1978; Cox 2001; Petney & Andrews 1998), and the neglected tropical diseases (NTDs) in particular are said to overlap and to cluster within poor communities in endemic areas (Aagaard-Hansen & Chaignat 2010). To date, studies reporting on multiple infections with gastrointestinal parasites have tended to do so at the individual level (Bisanzio et al. 2014; Hürlimann et al. 2014; Garbossa et al. 2013; Nguhiu et al. 2009; Mupfasoni et al. 2009; Midzi et al. 2008; Raso et al. 2004; Pullan et al. 2008). However, many of the risk factors that have been identified for mono- and co-infection in these studies, such as crowding, socioeconomic status, and environmental conditions, operate at the household-level. Moreover, within-household transmission for infectious agents is rarely complete (i.e. the intra-cluster correlation is generally a lot less then one, as we observed in Chapter 4) and therefore pathogen species richness at the individual level may tend to be lower than that at the aggregate level of the household. Quantifying household-level polyparasitism (or pathogen species richness), and exploring the factors that influence it, could assist with the design of integrated disease control programmes for multiple pathogens that can be targeted at the household-level.

In the previous chapter, we described how the infection status of a household could be modelled using zero-inflated binomial (ZIB) regression. This approach provided some account for imperfect detection within sampled households, and allowed both the probability of household infection and the probability of detection given infection to be conditioned by covariates. We extend our application of these models in this chapter and describe household-level infection for six neglected helminth parasites, *viz.* hookworm, *Ascaris*, *Trichuris*, *Strongyloides stercoralis*, *Taenia solium* and *Schistosoma mansoni*. We use ZIB models to derive an adjusted estimate of household-level prevalence for each species, as well as adjusted estimates of household-level species richness (or household polyparasitism). We focus in particular on examining how similarity in response to ecological conditions might result in household-level co-infection. To this end, we apply a multi-species extension of the

site occupancy modelling approach described in Chapter 6. These hierarchical 'community models' provide a framework that allows the estimation of the factors that influence occupancy status of single species and, importantly, the average response of a 'community' of co-occurring species (Dorazio & Royle 2005; Dorazio et al. 2006).

7.1.1. Background on the statistical methodology

It would be possible to perform multiple single-species site occupancy (ZIB) models for each of the neglected helminths under study, as we did for the STH species in Chapter 4 (and as others have done for non-parasitic species (Ferrier & Guisan 2006)). Such an approach would allow us to derive species-specific estimates of the probability of occupancy whilst controlling for detection, and the influence of covariates on both of those processes. Whilst the STH were all reasonably common, this approach might lead to unstable estimates for the rare species, or those that were difficult to detect (Zipkin et al. 2009; Burton et al. 2012). This would be of particular concern for *Strongyloides*, which was only detected in around 3% of individuals. Moreover, the approach is not particularly parsimonious: the number of parameters to estimate increases as the number of species within the community under study increases (Royle & Dorazio 2008).

Community models provide a means for dealing with issues of parsimony as well as inference for rare species, and do so by treating each species within a community as a random effect within a hierarchical framework (Dorazio & Royle 2005; Dorazio et al. 2006). Whilst the probability of both detection and occupancy can be expected to vary from species to species, community models are based on the assumption that these probabilities come from a common distribution of responses (Dorazio et al. 2011). The use of this common distribution to describe species-specific effects will tend to lead to increased precision and accuracy in parameter estimates (so called "Bayesian shrinkage" (Link & Sauer 1996)), particularly for those species that were rarely observed (DeWan & Zipkin 2010; Zipkin et al. 2009). Moreover, since parameter estimates for each species are dependent (in that they come from a common distribution of responses defined using the same hyperparameters), the approach also allows the formal estimation of characteristics of the observed community, including the mean 'community' response to ecological conditions.

Community models have the same conditional structure as single-species site occupancy models ((MacKenzie et al. 2002) and described in Chapter 6) but with an additional hierarchical layer to describe the random species effect (Kery & Royle 2008):

State process: $$z_{ik} \sim Bernoulli(\psi_k)$$

Observation process: $$Y_{ijk} \mid z_{ik} \sim Bernoulli(z_{ik} p_k)$$

Species heterogeneity: $$logit(\psi_{ik}) = alpha.psi_k + beta.psi_k * x_i \ldots.$$

$$logit(p_{ijk}) = alpha.p_k + beta.p_k * x_{ij} \ldots.$$

$$alpha.psi_k \sim Normal(\mu_{alpha.psi}, {\sigma^2}_{alpha.psi})$$

$$beta.psi_k \sim Normal(\mu_{beta.psi}, {\sigma^2}_{beta.psi})$$

$$alpha.p_k \sim Normal(\mu_{alpha.p}, {\sigma^2}_{alpha.p})$$

$$beta.p_k \sim Normal(\mu_{beta.p}, {\sigma^2}_{beta.p})$$

Hence, we consider that the presence or absence of species *k* at site *i* as a Bernoulli process with probability ψ_k. The observation process (detection) for species *k* at site *i* in survey *j* is also a Bernoulli process with probability p_k and is dependent on the occupancy status of the site. Both ψ_k and p_k are conditioned on covariates (x_i and x_{ij}) via a logistic regression in which (and in contrast to the single-species site occupancy models) the intercept and slope for each coefficient is defined by a random (normal) distribution with community-level hyperparameters (mean and standard deviation). Summarising these hyperparameters gives an indication of the community response.

Hierarchical community models have been used widely used in ecology (Iknayan et al. 2014), including for birds (Zipkin et al. 2009; Zipkin et al. 2010), carnivores (Burton et al. 2012), insects (Dorazio et al. 2006), and fish (MacNeill et al. 2008). To our knowledge, this is the first application to parasitic species, and certainly to organisms of medical importance.

7.1.2. Background on the life cycles of the NTDs under study

For standard applications of site occupancy models, there should be a theoretically infinite pool of potential detections at each site (MacKenzie et al. 2006) and therefore in the single-species models described chapter 6, we interpreted ψ to represent the likelihood that each STH species occurred in the household environment, rather than in at least one of the people we sampled. Fundamental to this assumption was that each person within a household (and therefore each person we sample) has a non-zero probability of infection given that the household in which they reside is occupied ('infected') by the parasite of interest.

We believe we can make the same general assumption on the interpretation of ψ for the other helminth parasites considered in this chapter (*Strongyloides stercoralis*, *Taenia solium* and *Schistosoma mansoni*). However, there are some important differences in the life-cycle

of each species that need to borne in mind when assessing the appropriateness of these assumptions.

i. STH and Strongyloides

The simplified life-cycle for the three STHs (*Ascaris*, *Trichuris* and hookworm) is shown in Figure 7.1, and involves the shedding of eggs into the environment via faeces, the development of eggs to infective forms (with rate and success dependent on environmental conditions), and exposure and infection of susceptible individuals via contact with contaminated soil. *Strongyloides stercoralis* has a similar lifecycle, but, and in contrast to the STH, individuals may also be auto-infected by resident worms within the GI tact (Figure 7.2). Infections therefore tend to be highly persistent (Mansfield et al. 1996), and each adult worm in the GI tract is not necessarily the result of a single environmental exposure. This is not necessarily a problem for our interpretation of occupancy (since environmental transmission still occurs), but infection intensity (and therefore the probability of detection on the basis of a single faecal sample) may be related to individual level factors such as immunological state and nutrition (Iriemenam et al. 2010) to a greater extent than for the STH.

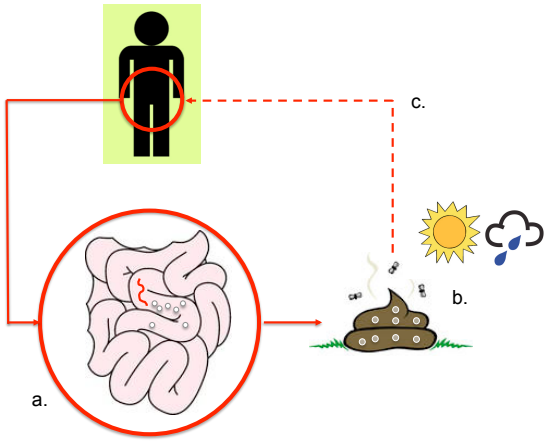


Figure 7.1. Simplified life cycle of the soil-transmitted helminths. a) Adult worms reside in small intestine where they produce eggs; b. Eggs shed in faeces and mature under appropriate environmental conditions; c. Susceptible individuals come into contact with infective eggs (*Ascaris* and *Trichuris*) or larvae (hookworm) and the cycle continues.

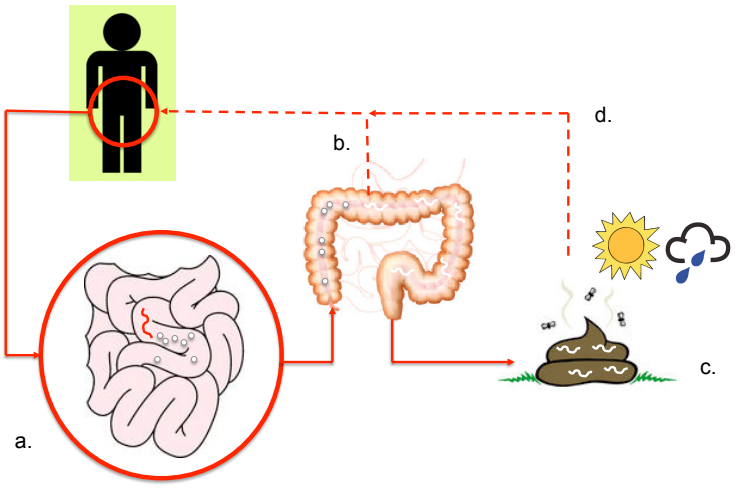


Figure 7.2. Simplified life cycle of *Strongyloides stercoralis*. a) Adult worms reside in small intestine where they produce eggs; b. Eggs may mature to larvae as they pass through the large intestine. These can penetrate the intestinal wall and the cycle continues (autoinfection); c. Larvae that have not gone through autoinfection are passed in faeces where they are subject to environmental conditions; d. Some larvae become parasitic, and infect individuals by skin penetration and the cycle continues.

ii. Schistosoma mansoni

The life cycle of *S. mansoni* depends on water snails as an intermediate host (Figure 7.3). Our definition of household occupancy therefore needs to be extended to include a household's water supply, and will be influenced to a large extent by the ecological factors that determine the presence or absence of snails of the *Biomphalaria* genus. These factors may be substantially different from those that influence occupancy with the other neglected helminths.

As is the case for the STH, diagnosis of *S. mansoni* is based on the identification of eggs in faeces, and can therefore be expected to be confounded by factors related to intensity of infection (Krauth et al. 2012).

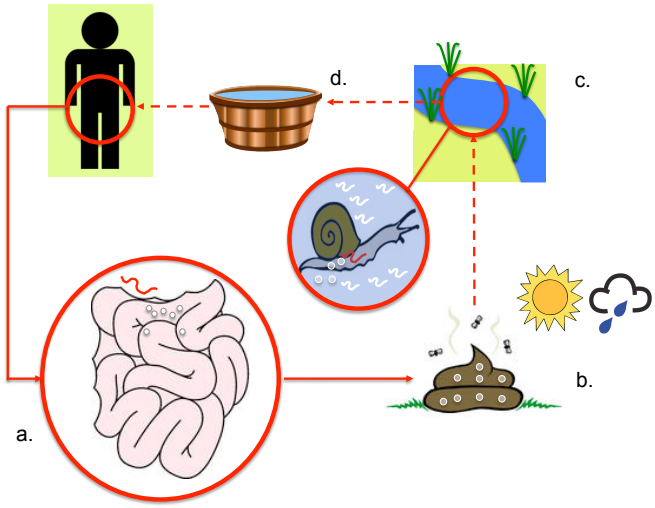


Figure 7.3. Simplified life cycle of the *Schistosoma mansoni*. a) Adult worms reside in blood vessels around the small intestine (or rarely the bladder) where they produce eggs; b. Eggs shed in faeces (rarely urine) enter surface water (directly or via run-off); c. Eggs hatch to miracidia and infect snails of *Biomphalaria* genus and develop to infective larvae. Infective larvae escape snail; d. People become infected by coming into contact with water contaminated by larvae during bathing etc.

iii. Cysticercosis

Cysticercosis occurs due to the presence of larvae of *T. solium* in the tissues of an infected person, rather than the presence of adult worms in and around the GI tract. Detection is therefore based on serological testing rather than faecal examination. However, and consistent with the STH and *S. mansoni*, each parasitic form (cyst) present within an infected individual is the result of a single exposure from the environment (or through direct faecal-oral transmission if the person has an adult tapeworm) (Figure 7.4). The sensitivity of the HP10 ELISA, the antigen capture test used in this study, has been shown to vary according to the number of cysts present in an infected individual: the greater the number of cysts, the greater the diagnostic sensitivity (Garcia et al. 2000; Fleury et al. 2003; Lescano et al. 2009). Detection can therefore be considered somewhat analogous to the intensity related process already described for the other helminth infections.

An important distinction from the other helminths is that an individual with cycticercosis is entirely non-infectious, and represents a dead-end in the life cycle for *Taenia solium*. In order for a household to be considered occupied (such that everyone we sample within it has a non-zero probability of infection), it is therefore necessary for there to be a tapeworm

carrier in the household or within its extended environment (which is not necessarily limited to household residents and could include, amongst other scenarios, regular visitors or neighbours who defecate in the household's fields). Detection, by contrast, is based on the presence of cysterci in household members.

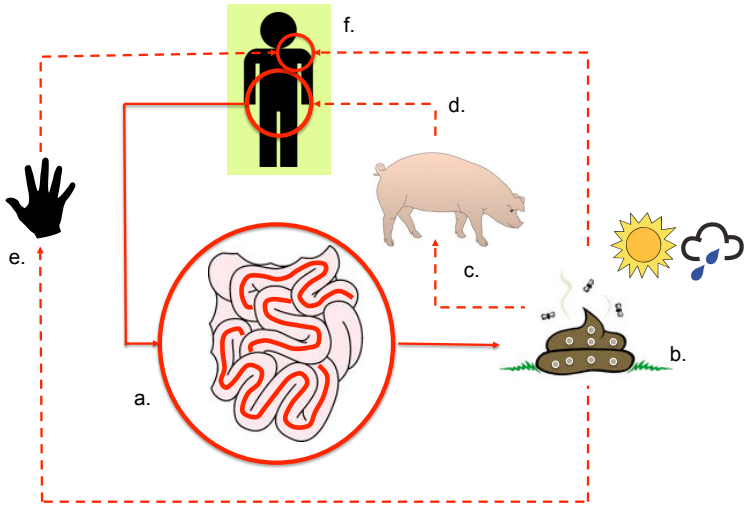


Figure 7.4. Simplified life cycle of the *Taenia solium*. a) Adult worm resides in small intestine where it produces eggs. b) Eggs shed in faeces; c) Eggs may be ingested by scavenging pigs: ingested eggs hatch and larvae enter circulation and encyst in tissues (porcine cysticercosis); d) Infective cysts consumed in inadequately cooked pork may develop into adult worm in small intestine (taeniasis): e) Eggs do not require a period in environment to become infective and autoinfection can occur due to inadequate personal hygiene; f) Following environmental exposure or autoinfection, ingested eggs hatch and larvae enter circulation and encyst in tissues (human cysticercosis).

Given that one of our primary aims for this chapter was to identify common responses to external conditions amongst this group of medically important parasites, we do not see this heterogeneity as a major problem. However, some of the potential consequences for modelling a 'community' of diverse parasites (and by community we refer to a set of species occurring in the same area) are explored in the discussion.

7.2. Methods

We defined household occupancy for each of the six helminth species under study on the basis of single samples collected from multiple household residents.

7.2.1. Model structure

As is the case for single species site occupancy models, hierarchical community models allow the inclusion of covariates in the logistic regression equation describing the probability of occupancy and that for detection given occupancy.

Covariates on household occupancy

A key issue for the control of environmentally transmitted helminth infections, and indeed for most infectious diseases, is that the risk of household infection will be influenced by the disease status and practices of the households that surround it (Esrey et al. 1991; Appleton et al. 2009; Schmidlin et al. 2013; Chongsuvivatwong et al. 1996; Feachem et al. 1983; Asaolu & Ofoezie 2003). The local prevalence of infection, and therefore probability of household occupancy, might be expected to be influenced by a range of social and environmental factors. These factors, and those related to household susceptibility to infection, are summarised in Figure 7.5A and described in further detail below.

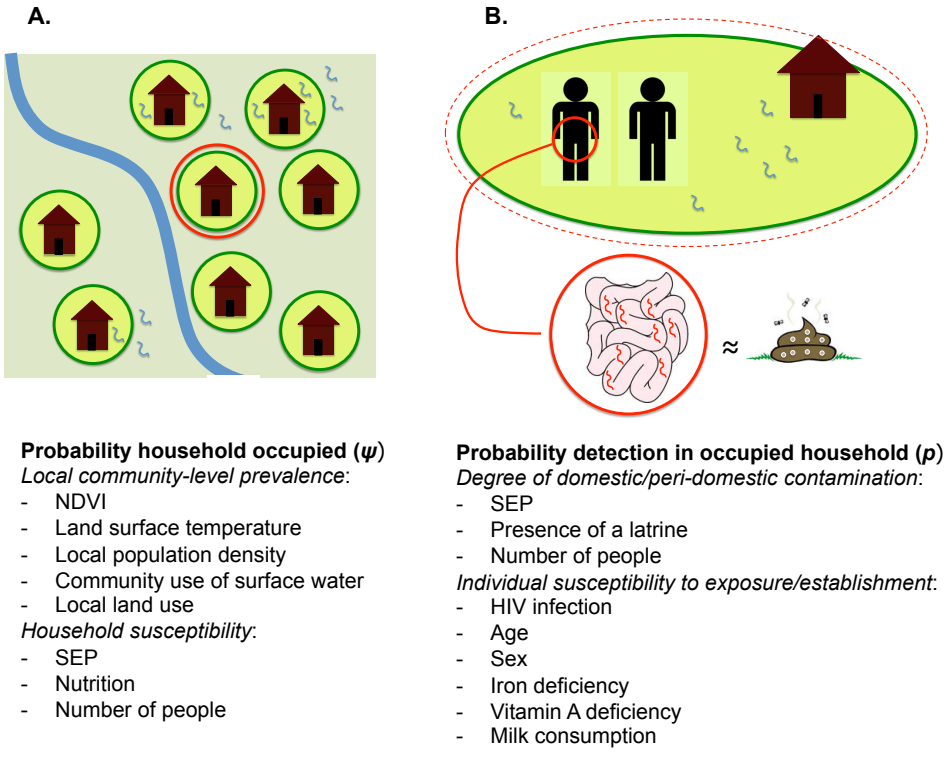


Figure 7.5. Covariates on occupancy (A) and detection (B). We seek to determine the occupancy status of a household with a parasite given that it occurs in our study site (A). Occupancy status is defined on the basis of detection in samples from individuals who may or may not be infected if a household is occupied (B).

i. Environmental conditions

Local vegetation density and/or land surface temperature could be expected to influence the environmental stages of each of the parasites of interest (Pawlowski 2002; Brooker et al 2006b; Xue et al. 2011; Khieu et al. 2014). We used the average NDVI and maximum LST extracted at the household location from Fourier processed data ((Scharlemann et al. 2008) and used in chapters 5 and 6). Land cover, and particularly grass area, has been linked to community level socioeconomic status in western Kenya (Okwi et al. 2007). We used the land cover data from Wardrop *in prep* (described in chapter 5) and defined the percentage of the local area that was under grass and crop land within a buffer of 1km of each household.

ii. Population density

Increased population density may increase parasite transmission success (Poulin 2004), and therefore local community prevalence, or conditions in areas with greater population density (e.g. that are more urban) may reduce suitability for environmental life stages, or impact upon individual exposure (Phiri et al. 2000; Pullan, Kabatereine, et al. 2010b). We used the bespoke index of habitation density (first described in Chapter 5 and also used in Chapter 6) to represent local population density.

iii. Community use of surface water

Community water supply might be expected to directly influence community prevalence (as a result of use of water sources with high levels of parasite contamination) (Steinmann et al. 2006; Pullan et al. 2014a; Esrey et al. 1991; Soares Magalhães et al. 2011a; Vilhena et al. 1999), or may provide an indication of the level of infrastructural investment in the wider community. We derived the proportion of households per sub-location (the smallest administrative unit in Kenya) using surface water (rivers, lakes, ponds) for their domestic needs from the 2009 Kenya census (Open Kenya 2014). Equivalent data on latrine use (and other socioeconomic indicators) were not available at an equivalent resolution from open source data at the time of the study.

iv. Household SEP and nutrition

Household socioeconomic position (SEP) could be expected to influence vulnerability to a range of infectious diseases (Bates et al. 2004b), including the neglected helminths (King 2010; Gazzinelli et al. 2012). We defined SEP using the composite index from multiple

factor analysis (MFA) described in Chapter 5. A household's collective nutrition status might also impact upon its vulnerability to infectious disease (Bates et al. 2004b). Several studies have shown that cow ownership influences household nutrition (Ahmed et al. 2000; Mullins et al. 1996), particularly that of children (Hoddinott et al. 2014). We therefore included the presence or absence of a cow as a covariate on household occupancy with each parasite.

v. Household size

It could be expected that the opportunity for multiple parasite species to occur in the same household might increase as household size increases (and indeed, the species-area relationship, the principal that the number of species in an area increases with its size, is fundamental in ecology (Lawton 1999)).

Accounting for spatial autocorrelation

In chapter 4, we identified evidence for spatial dependency for all of the infections under study, particularly the STH and *S. mansoni,* and we might therefore expect residual spatial autocorrelation to be an issue. In an attempt to account for the possible effects of (unexplained) autocorrelation in household infection with each parasite, we included an auto-covariate term in the logistic regression equation describing occupancy. This autologistic extension (Augustin et al. 1996; Royle & Dorazio 2008), considers a neighbourhood of pre-specified dimensions around each surveyed household, and allows the presence or absence of the species of interest in all households within the specified area to influence (together with the effects of modelled covariates) the probability of occupancy.

Study households were selected from the study area at random (but stratified by sublocation, see chapter 3) and the resulting density was quite variable. As a result, there was neighbourhood missingness (i.e. zero neighbours) for some households up to a radius of 5 km, whilst others had more than 40 neighbours at this distance. We therefore considered a neighbourhood of pre-specified number around each site rather than pre-specified area, and allowed a household's (unobserved) infection status to be influenced by that of its 5 closest neighbours, with the degree of influence defined using a distance decay function. Further details and specification of the autocovariate in R language are given in the Appendix.

The general model for the occupancy status of species *k* at household *i* was therefore:

$$\text{logit}\,(\psi_{ki}) = \alpha_k + \beta1_k\,\text{NDVI}_i + \beta2_k\,\text{LST}_i + \beta3_k\,\text{Grass area}_i + \beta4_k\,\text{Pop. density}_i + \beta5_k\,\text{Surface water}_i + \beta6_k\,\text{SEP}_i + \beta7_k\,\text{Cow}_i + \beta8_k\,\text{Number people}_i + \delta_k\,\text{autocovariate}_i$$

Covariates on the probability of detection

For all of the infections of interest, we assume that the probability of detection in individuals living in an occupied (infected) household will be a function of the intensity of infection and, in the case of STH, *S. stercoralis* and *S. mansoni*, the fecundity of resident parasites. Factors that influence infection intensity are likely to include the degree of contamination of the domestic and peri-domestic environment, as well as the susceptibility of individuals following exposure (and to autoinfection in the case of *Strongyloide*s) (Figure 7.5B). The degree of domestic and peri-domestic contamination will be linked to factors such as sanitation, education, and building quality, and therefore to the SEP of the household (Lim et al. 2009). We therefore included household SEP as a predictor of detection in addition to occupancy. Given the particular importance of sanitation (and environmental contamination) for all of the infections of interest, and acknowledging its potential mediating effect on SEP, we also included latrine ownership as a separate effect. Age and sex can be considered to influence individual exposure, as well as susceptibility to infection. HIV infection status, iron and Vitamin A deficiency and the frequency of milk consumption may also influence susceptibility to infection, and potentially the fecundity of resident parasites. The methods used to classify iron and vitamin A deficiency were described in Chapter 3 (or briefly, they involved the adjustment of ferritin and retinol binding protein concentrations on the basis of indicators of inflammation (Erhardt et al. 2004)). Reported frequency of milk consumption was dichotomised into daily and less than daily.

The probability of detection of species *k* in household *i* on the basis of the evaluation of a single sample from person *j* was therefore:

$$\text{logit}\,(p_{kij}) = \alpha_k + \beta1_k\,\text{age}_{ij} + \beta2_k\,\text{sex}_{ij} + \beta3_k\,\text{HIV}_{ij} + \beta4_k\text{Iron deficiency}_{ij} + \beta5_k\,\text{Vit. A deficiency}_{ij} + \beta6_k\,\text{Milk consumption}_{ij} + \beta7_k\,\text{SEP}_{ij} + \beta8_k\,\text{Number\_people}_{ij} + \beta9_k\,\text{Latrine}_{ij}$$

Model extensions

A basic principle of ecological theory is that common species tend to be both more widespread and more locally abundant than rare species (Brown 1984), and therefore the probability of occupancy and detection (which is often a function of abundance (Royle 2006)) may be linked (Zipkin et al. 2009). Moreover, and in terms of parasite transmission, households in which a parasite is most 'abundant' (in the environment and in infected individuals) and in which there are high levels of within-household transmission are likely to remain as occupied for longer than those in which there is little transmission (again, linking occupancy with detection). We therefore modelled the covariance between ψ and p by assuming a parametric form for the joint distribution of both probabilities for each species, so that $[alpha.psi_k, alpha.p_k|\Sigma] \sim N(0, \Sigma)$ with the 2x2 matrix Σ specified by two variances ($\sigma^2_{alpha.psi}, \sigma^2_{alpha.p}$) and a covariance ($\sigma_{alpha.psi,alpha.p}$) (Kery & Royle 2008).

Model specification

Models were fit in WinBUGS via the *R2WinBUGS* package (Sturtz et al. 2005). Model convergence was confirmed by visual assessment of MCMC chains and on the basis of the Rhat statistic (<1.1 suggesting convergence (Gelman et al. 2013)). The model were fit with uninformative priors, following (Kery & Royle 2008), and inference was based on 3 chains which were allowed to run for at least 70,000 iterations after a burn-in of 20,000. The R code used to define the community model is given in the Appendix.

To assist model convergence, all continuous predictors (of both detection and occupancy) were scaled by subtracting the mean and dividing by one standard deviation.

Assessing variable importance

Model selection procedures are less developed for Bayesian than frequentist approaches (O'Hara & Sillanpää 2009), and are somewhat limited for complex hierarchical models where information criteria such as DIC (as an indicator of model fit penalised by complexity) are unreliable (Royle & Dorazio 2008). We fit and present the full model (with all covariates considered a priori to be important or interesting) and did not attempt a variable selection procedure. The posterior distributions of the effect of all covariates at both the community and individual species level were examined, and effects were considered to be 'significant' when the 95% credible intervals did not include zero.

Model diagnostics

We assessed model fit using a posterior predictive checking approach (Gelman et al. 1996). Calculation of the so-called 'Bayesian p-value' involves simulating data sets under the model and comparing the discrepancy between predicted and observed data (Gelman & Hill 2006). The proportion of discrepancy measures that are greater than that for the observed data (over multiple draws during MCMC runs) provides the 'p-value'. Values close to 0 or 1 are considered suspicious, whilst those around 50% represent good agreement (Kery & Schaub 2012).

Household parasite species richness and its relationship with HIV

In Chapter 4 we described a possible relationship between the spatial distribution of HIV and elevated risk for multiple infections, and in Chapter 5 and 6 the count of HIV positive individuals within a household was a significant predictor of individual risk for *Trichuris* and *Ascaris* infection (or, more correctly, detection). To further explore the potential impacts of household-level parasite richness on individual-level risk of HIV, we used the average estimated parasite species richness (defined as the sum of predicted occupancy states for each parasite in each household, i.e. $\sum z_{ik}$) as a predictor of individual infection. A multivariable model was built to explore the relationship between individual HIV risk and household parasite count, and which included control for age, sex and the number of people in the household, together with a random effect at the household level to account for the correlated nature of observations. Improvement in fit following inclusion of a restricted cubic spline (RCS) with 3 or 4 knots on the continuous predictors of age and number of people was assessed on the basis of AIC.

We also examined (using the logistic regression approach described above) the relationship between individual HIV infection and the observed household-level parasite count (species richness), as well as that derived from a null community model in which detection was not conditioned on covariates, and only the autocovariate term was included in the model describing occupancy.

In all cases, residual spatial autocorrelation was assessed using Moran's I on household-level residuals (as described in chapter 5). The mixed effects model was fit using the *lme4* package in R (Bates et al. 2014), restricted cubic splines were included via the *rms* package (Harrell 2006), and *Moran's I* was assessed using the *ape* package (Paradis et al. 2004) .

7.3. Results

The full community model, containing all species of interest and the spatial autocovariate term, resulted in a Bayesian p-value of 0.564, suggesting no evidence for a lack of fit.

7.3.1. Household prevalence

The proportion of sites considered occupied (i.e. the household prevalence) on the basis of the community model was higher than the observed for each infection, and the difference was quite substantial for *Strongyloides* (Table 7.1). Indeed whilst this parasite was observed in the fewest households, we predict that it may be the second most prevalent (in terms of number of occupied households) helminth parasite in our study area, following hookworm. However, it is important to note that the posterior distribution is quite wide and there is therefore considerable uncertainty in this estimate.

Unexpectedly, species occupancy and detection were negatively correlated (rho=-0.30), although the posterior distribution was extremely wide (95% CI -0.90, 0.62).

Table 7.1. Observed and adjusted household prevalence estimates.

Parasite	Observed (95% CI)	Adjusted (95% CI)
Ascaris	23.8 (19.9-28.3)	39.0 (32.8-47.1)
Hookworm	69.6 (64.8-73.9)	79.2 (76.0-82.8)
Trichuris	30.4 (26.0-35.2)	50.1 (42.6-58.6)
Taenia solium	17.8 (14.3-21.9)	27.2 (22.8-33.3)
Strongyloides stercoralis	13.4 (10.3-17.1)	75.6 (45.4-97.5)
Schistosoma mansoni	19.0 (15.4-23.2)	24.7 (21.8-28.4)

7.3.2. Individual species and community-level effects

Occupancy

The 'average' effect of covariates on household occupancy for the community of helminths under study is shown in Figure 7.6, and the species-specific effects in Figure 7.7 (actual values are given in Appendix 7.3). Only household size appeared to have a consistent effect on the probability of occupancy for each species (Figure 7.7), and the posterior mean distribution for its effect at the community-level did not include zero (Figure 7.6). There was also some indication that the proportion of the local area under grass and crop land had a negative effect on occupancy for all species excluding *Taenia*, although the posterior

distribution at the community level (Figure 7.6), and that for all individual species (Figure 7.7) included zero across much of its range.

The NDVI had a significant effect on the probability of household occupancy with *S. mansoni*, with households in areas with lower vegetation density appearing to have lower risk of household-level infection. A similar effect could be observed for maximum LST, although the 95% credibility interval overlapped zero to a small extent. There was also some indication that household occupancy with *T. solium* was less likely in areas with high vegetation density, although again, the 95% CI overlaps zero *mansoni*. In the previous chapter, we observed that higher NDVI was an important predictor of household infection with hookworm. We observe a similar effect here, but there was less evidence to support it (where we include some control for spatial dependency in observations) (Figure 7.7).

There was no obvious effect of population density, SEP, or ownership of a cow on household occupancy for any of the individual parasites. An increasing proportion of households using surface water within a sub-location appeared to reduce the risk of hookworm infection, whilst there was weak evidence of a positive effect on the household risk of *Trichuris* and *Schistosoma* infection.

A strong and significant spatial effect was observed for all species, particularly for *Trichuris* and *T. solium*. A comparison of the outputs from the spatial model was made with the same model without the autocovariate term (i.e. a non-spatial model) and is presented in Appendix 7. The sign and magnitude of all effects (and therefore their interpretation) is generally the same, but a larger number of significant effects are seen at the species-level in the non-spatial model (and therefore the autocovariate term tends to decrease the effect size of the other predictors or to increase the uncertainty).

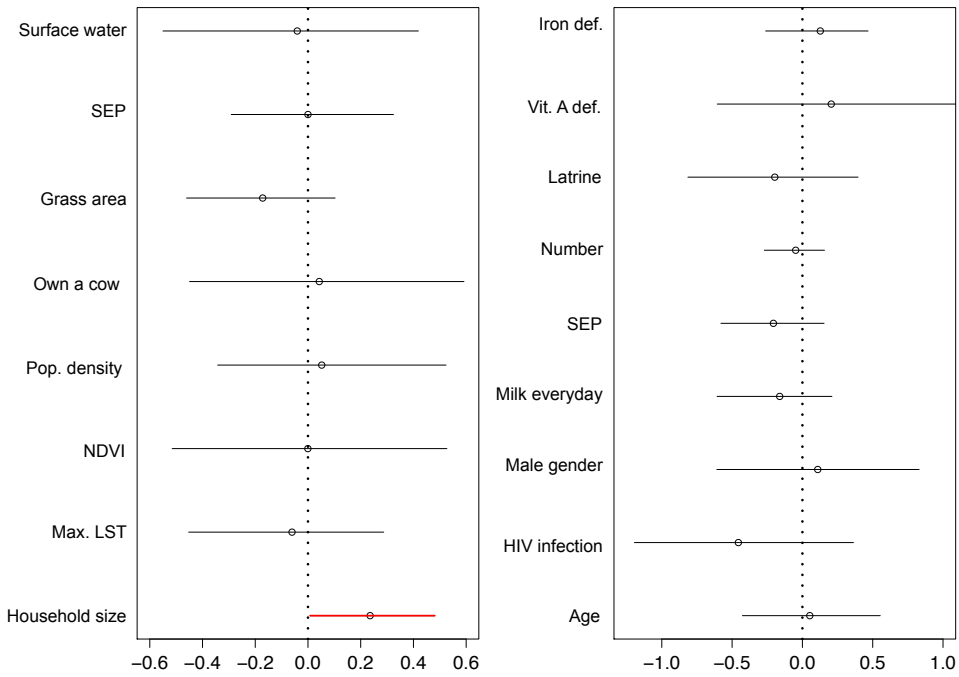


Figure 7.6. Community-level effects of covariates on the probability of occupancy (left) and detection (right). Points show posterior mean with 95% credibility intervals (red lines do not include zero).

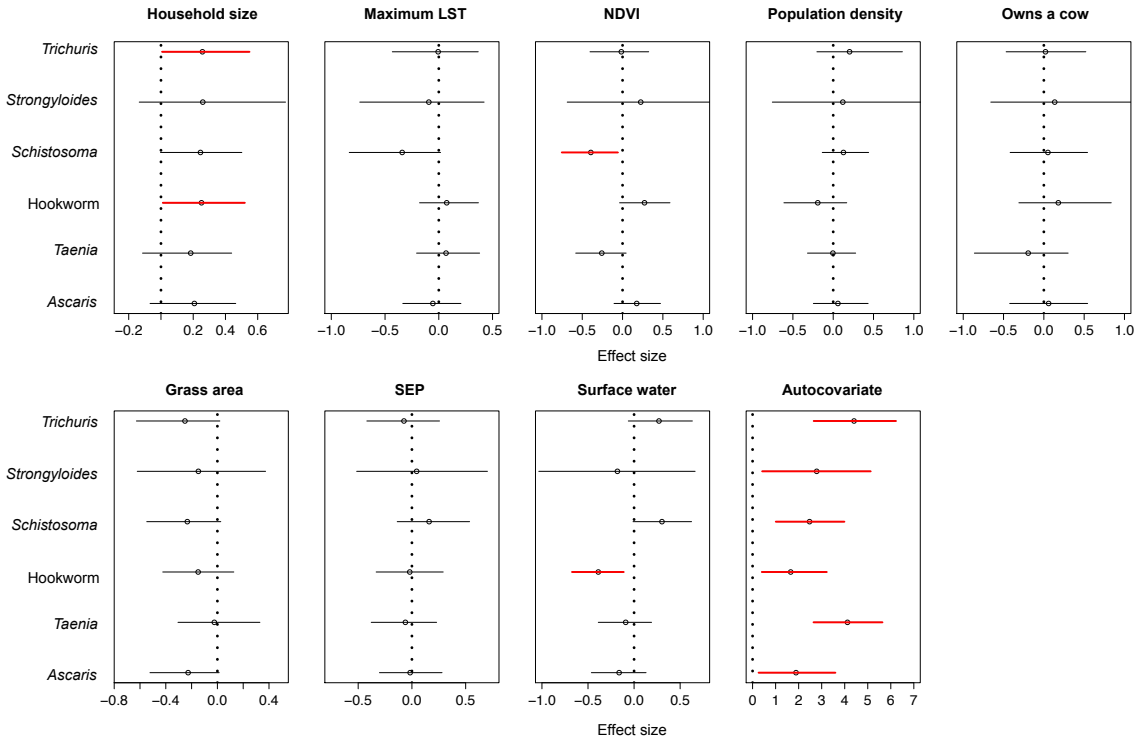


Figure 7.7. Species-level effects of covariates on the probability of occupancy. Points show posterior mean with 95% credibility intervals (red lines do not include zero).

Detection

There was no evidence of a consistent group effect for any of the covariates on detection (Figure 7.6). A similar (positive) response for iron deficiency was observed for all species, however in no case was this significant and 95% confidence intervals broadly overlapped zero (Figure 7.8). Other than for *T. solium*, HIV infection resulted in a reduced probability of detection for all parasites, and this effect was significant in the case of hookworm and *Trichuris*. There were quite heterogeneous effects of age and sex for each infection, although minimal effect of age was observed on the probability of detection for either *S. mansoni* or *T. solium*.

Increasing household SEP had a negative effect on the probability of detection for *S. mansoni*, hookworm and *Ascaris*, but no effect on *T. solium* and a positive (but insignificant) effect on *Trichuris* infection. Even with control for SEP, the presence of a latrine (of any quality) in the household resulted in a significant reduction in the probability of detection of *S. mansoni* in an occupied household. Consuming milk every day also reduced the probability of detection for this parasite.

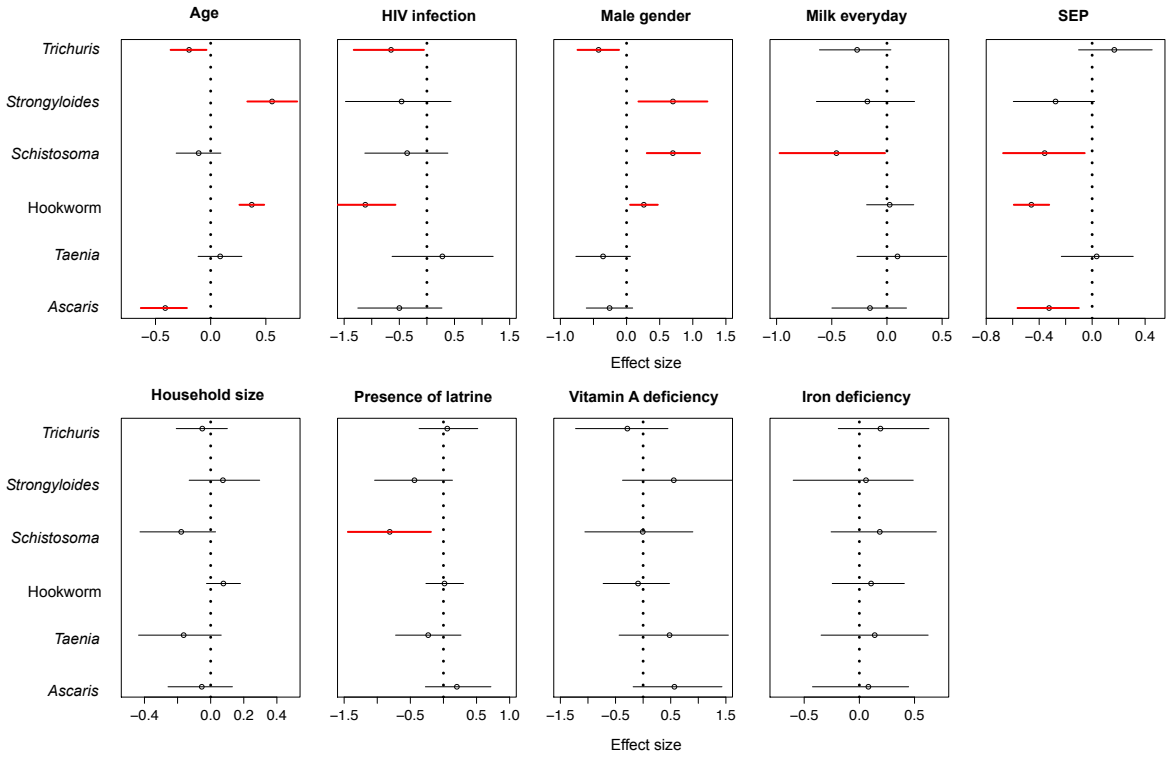


Figure 7.8. Species-level effects of covariates on the probability of detection. Points show posterior mean, lines show range of 95% credibility intervals (red lines do not include zero).

7.3.3. Household species richness and its relationship with HIV

The average number of helminth species observed in each household (out of a maximum of 6) was 1.74, and ranged from 0 to 5. The predicted average species count was 3.0, with a range from 0.94 to 5.96. The comparison of observed and adjusted species counts at the household level is plotted in Figure 7.9.

This plot also includes the count of HIV positive people per household, and suggests a possible trend for increasing number of HIV infected individuals in those households that have the highest average parasite counts. This is supported at the individual level, where a significant positive relationship was observed between risk of HIV infection and household parasite count (species richness) (Table 7.3). A similar effect was observed for the null model and for the observed data, although the magnitude was less substantial in the case of the latter, and there was much less evidence to support it.

There was strong evidence for residual spatial autocorrelation in all of these reasonably simple models (Moran's I = 0.03, p = <0.001 in the case of the model using predicted parasite counts from the full community model).

Age and the number of people living in a household both had non-linear relationships with individual HIV infection, and in both cases the best fitting models included a spline with 3 knots. The effect of both predictors in the model with observed parasite species counts is shown in Figure 7.9, and this effect was basically the same in all three models (and therefore is not shown for each).

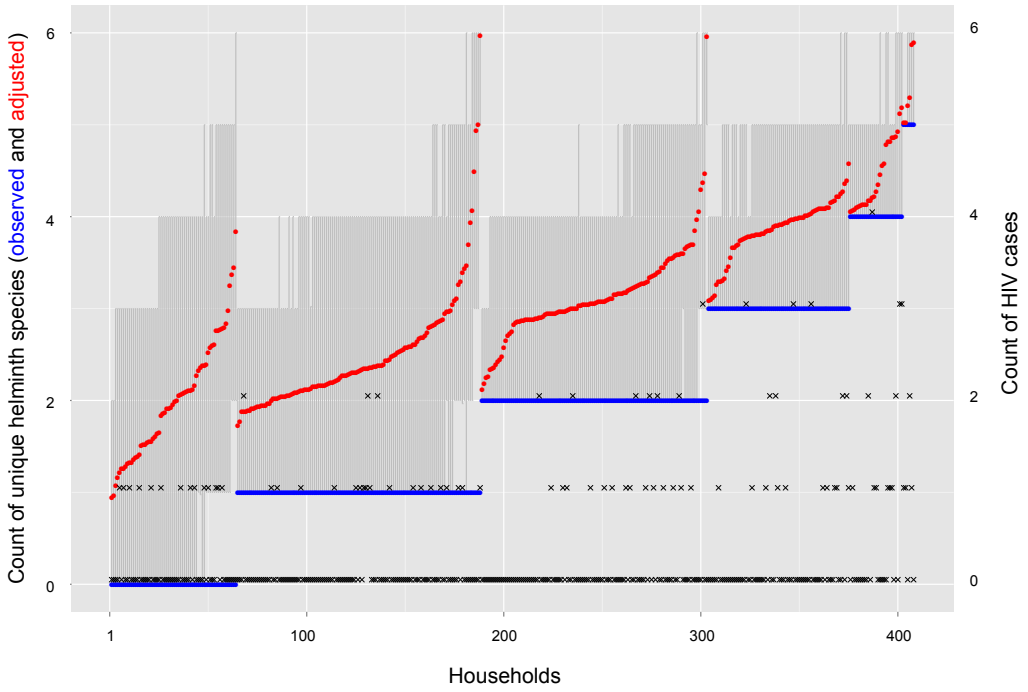


Figure 7.9. The observed count of helminth species per household (blue), adjusted count per household (red) with 95% credibility intervals. The count of HIV positive individuals per household is also shown (black x's). Households have been ordered based on both observed and adjusted counts.

Table 7.2. Outputs from multivariable logistic regression comparing the effect of the household parasite species count on individual risk of HIV infection (observed or unadjusted counts; counts derived from the full model; and counts derived from the null model)

	Observed		Modelled (Full)		Modelled (Null)	
Predictor	**OR (95% CI)**	**p-value***	**OR (95% CI)**	**p-value***	**OR (95% CI)**	**p-value***
Parasite count	1.26 (0.98, 1.62)	0.060	1.60 (1.19, 2.15)	0.0011	1.50 (1.13, 2.00)	0.0036
Age	*see Figure 7.10A*	<0.001	*see Figure 7.10A*	<0.001	*see Figure 7.10A*	<0.001
Male gender	0.52 (0.32, 0.87)	0.005	0.53 (0.32, 0.87)	0.005	0.53 (0.32, 0.87)	0.005
No. of people	*see Figure 7.10B*	0.006	*see Figure 7.10B*	0.006	*see Figure 7.10B*	0.01

* Estimated on the basis of likelihood ratio test.

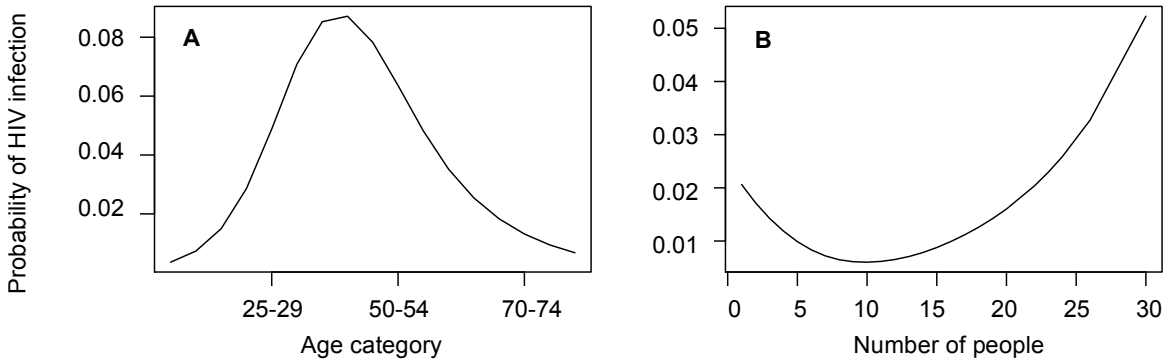


Figure 7.10. Non linear relationship between age (A) and count of people per household (B) and individual risk of HIV infection.

7.4. Discussion

There is growing interest in the epidemiology of polyparasitism, and particularly in the effects of environmental and social conditions on the risk of multiple infections in multi-endemic communities (Gazzinelli et al. 2012). For the helminths, most studies have focused on individual co-infection, or the "wormy individual" with few reporting on the likely phenomenon of a "wormy household". In this chapter, we have explored that phenomenon, and particularly the importance of the shared household-level risk factors that might influence it. Through the use of a hierarchical modelling approach that treated species as a random effect, we have been able to formally assess the shared response of a group of medically important helminths to a range of social and environmental conditions. Furthermore, and using zero-inflated binomial regression, we have modelled this household-level outcome whilst controlling for the conditions that may influence the probability of detecting infection in a household given that it is occupied by the parasite, or the sensitivity of our within-household sampling procedure.

Model outputs suggest that parasite responses are quite heterogeneous at the household-level, and only in the case of household size was a consistent group (or 'community') response observed. The species under study have quite different lifecycles (as we describe in section 7.1.2). Dorazio et al. (2011) caution against using community models to explore the effects of ecological conditions on very dissimilar species (giving the example of an unlikely community of mice and tigers), while Pacifici et al. (2014) specifically explored the impacts of species composition on inference. As might be expected, they found that homogenous responses amongst more similar species could be averaged out and lost when they were included in a community of dissimilar species with heterogeneous responses. An argument could therefore be made for limiting the analysis to the helminth species with the most similar life cycles (perhaps the STH and *Strongyloides*). However, since our primary interest was to identify common responses amongst the whole group of neglected parasites, the identification (or confirmation) of heterogeneous responses to a range of environmental conditions (as well as the homogenous response to household size) is of interest in itself.

A sensitivity analysis of the effects of excluding particular species may, however, be a useful and interesting next step. Indeed, the same analysis of a group of parasites that did not include *Taenia solium* might reveal a more substantial community response to the proportion of local area under grass and crops, the negative effect of which was reasonably consistent for the other parasites (but for which 95% credibility intervals were quite wide). Okwi et al. (2007) found that increasing grass area in a location (the second administrative unit in

Kenya) was associated with decreasing levels of poverty in Western province (in which most of our study area is located), which could explain the observed response amongst these neglected helminths. Household occupancy with a *Taenia solium* tapeworm is likely to be influenced by pig ownership and/or pork consumption practices: the land use factors (and other household level predictors) that influence these practices may well be very different to those that influence exposure to infective parasites via the environment..

We also sought to account for local poverty by including a covariate describing community use of surface water (at the sub-location level, the first administrative unit in Kenya). The only significant effect observed was a reduction in the probability of hookworm occupancy as the proportion of households within a sub-location using surface water increases. This effect is difficult to explain, and unmeasured confounding is perhaps quite likely. More intuitive (but non-significant) relationships were observed between *S. mansoni* and *Trichuris*, and these may be directly associated with increased exposure to contaminated water sources rather than (or in addition to) an indirect effect of infrastructural poverty, particularly in the case of *S. mansoni* (Fenwick 2006; Steinmann et al. 2006).

Kenya conducted a household census in 2009, which included the collection of indicators of household level poverty that have the potential to be aggregated to the community level (e.g. proportion of households living under the poverty line, education, access to sanitation). Whilst summarised data at a sufficiently high spatial resolution were not available for this study, they could potentially be used (following a successful application to Kenyan Office for National Statistics via the Kenya Open Data Initiative) to further explore the contextual effects of poverty on multiple infection risk.

In chapter 4, we observed substantial heterogeneity in the spatial distribution of the parasites explored here, and it is striking that the autocovariate terms for each species were highly significant. This suggests that there are environmental and social factors operating within the scale of our study area that influence household risk of infection that we have not adequately accounted for. This lack of account may be due to issues with the household level covariates we used. In particular, data describing land surface temperature and NDVI were at relatively (for the size of the study area) low resolution, and a stronger effect (and better account for between household spatial dependence) may be observed with more highly resolved data. In addition to these environmental effects, further analyses with better control for local levels of poverty, infrastructural investment and access to medical services may also help to explain the high levels of spatial dependence in each outcome.

As described in chapter 6, we consider household occupancy to be the presence of each

species in the environment of the household, and that people act as 'samplers' of this environment. This general consideration, and the subtle distinction of defining occupancy as a household-level process rather than being dependent on the small number of people we sample, allows us to meet the binomial assumptions of these models (MacKenzie et al. 2006). Our application therefore requires that all of the people we sample have a non-zero probability of infection. Clearly, people will also acquire infections outside the domestic and peri-domestic environment. Indeed, in the case of some of the parasites of interest, particularly hookworm (and perhaps therefore *Strongyloides*, since the two have similar life cycles (Khieu et al. 2014)), infection in the public 'domain' may occur more frequently than in the 'private' domain (Cairncross et al. 1996). This is not a problem, since the establishment of a patent infection in a resident of a household will result in infection (or occupancy) of that household. However, it is important that we are able to assume there is at least *some* onward transmission. Landscape genetics has revealed that most *Ascaris* infections are acquired in and around the home (Criscione et al. 2010), and this is also likely to be the case for *Trichuris* (Forrester et al. 1988). Transmission to other household members is also likely to be important for *T. solium* (Schantz et al. 1992; García et al. 2003), and clear effects of distance decay on the risk of transmission from a tapeworm carrier have been observed (Lescano et al. 2009), such that those people who live in a household with an individual with taeniasis are at high risk of cysticercosis. *Schistosoma* spp. infection also tends to cluster within households, largely because household members have similar contact with (contaminated) water sources (Bethony et al. 2001; Clennon et al. 2004) (which we include in our definition of household occupancy). There has been comparatively little work to explore within household transmission of *Strongyloides* but, like hookworm (which is also transmitted via skin penetration of larvae), is likely to occur (Lindo et al. 1995; Conway et al. 1995). Moreover, direct person to person transmission of *Strongyloides* has been reported (Czachor & Jonas 2000).

A possible scenario in which these assumptions may not be met might occur if a household member acquires a *Taenia solium* cystic infection when outside the household. These individuals would be entirely un-infectious, and therefore the household would not be "occupied" according to our definition of occupancy. We have not separated such false positives from 'true' occupancy, which would occur when all members of the household have some degree of exposure to *T. solium* eggs within the household and peri-domestic environment (e.g. due to a tapeworm carrier in the household). Methods to account for false positivity are available in the site occupancy literature (Royle & Link 2006). Moreover, although we did not include them in this chapter, we also collected data on taeniasis (i.e. the

presence of a tapeworm in the gastrointestinal tract), and these could potentially be used to separate these processes further. Richmond et al. (2010) described an extension of site occupancy (ZIB) models for describing occupancy of dependent species (with one dominant species impacting upon another, such as within a predator-prey relationship) and such models could potentially be explored for within-household taeniasis and cysticercosis dynamics, and particularly the dependence of the latter on the former.

Whilst there was little evidence of an effect of household SEP on the likelihood that a household is occupied by any of the parasites of interest, quite large effects were observed for detection, particularly for hookworm, *Ascaris*, *S. mansoni* and *Strongyloides*. This strongly suggests that whilst rich and poor households have an equal likelihood of being infected by each parasite, fewer people are infected, or have lower intensity infection (or resident parasites have lower fecundity) in richer households. Given the ubiquity of these parasites, and the likelihood of exposure in the public domain (Cairncross et al. 1996), this finding is perhaps not particularly surprising. Indeed, aggregated transmission is recognised to occur for many infectious diseases, such that most individuals and households experience low or zero levels, whilst a few experience very high levels (Woolhouse et al. 1997), and this is also a well-described feature of helminth transmission (Forrester et al. 1990; Chan et al. 1994b; Bradley & May 1978; Anderson & May 1992).

The adjusted estimates for household prevalence with *Strongyloides* were substantially higher than the observed data. Infection with this parasite was observed quite rarely at the individual level, but observed household prevalence was relatively high (3% vs. 13.4%). Few people were therefore observed to be infected within occupied households and therefore the average detection probability is quite low. Indeed, holding all predictors of detection at baseline and at the average value for the continuous outcomes, the probability of detection of *Strongyloides* in an individual in an occupied household was estimated as just 0.03. This probability is, of course, conditioned on covariates, and in an 80 year old male would increase to around 0.3. It is perhaps worth reemphasising (as we did in the previous chapter) that these values do not represent diagnostic sensitivity, since they include the probability that an individual is infected given that they live in an occupied household (in addition to the probability we detect the infection given that they are infected). With such a low detection probability, we can have very little certainty that a household is not infected if we do not observe infection, and given that there were no strong predictors of occupancy (excluding the presence of *Strongyloides* in neighbouring households and household size), most households *could* be occupied. The resulting estimate of household prevalence is therefore

very high, but has extremely wide credibility intervals. Several authors have suggested that the regional and global prevalence of *Strongyloides* is likely to be massively underestimated (Bisoffi et al. 2013; Montes et al. 2010) and much more work is therefore needed in the study area (and elsewhere) to determine the true prevalence of this infection. The results in this chapter (and those in chapter 4 where the adjusted household prevalence was estimated to be 61%) suggest that it could be quite high. Tests with higher sensitivity are available (e.g. Baermanns and Koga Agar plate) and are likely to result in a more precise estimate of household (and individual) prevalence (Glinz et al. 2010a).

In previous chapters, we described a significant negative relationship between HIV infection and individual infection with Trichuris and hookworm. We see the same relationship here, and a negative (but non-significant) effect for *Strongyloides*, *Schistosoma* and *Ascaris.* Only *Taenia* showed a (non significant) positive effect (and again, there might be more stronger evidence of a community-level effect of HIV on detection if *T. solium* was not included in the group of parasites). This negative relationship with HIV infection, has been widely described for the STH (as we describe in Chapter 6), as well as for *S. mansoni* (Karanja et al. 1997; Fontanet et al. 2000; Mwanakasale et al. 2003; Muok et al. 2013). Schaer et al. (2013) conducted a meta-analysis of the effect of HIV on *Strongyloides* and found that HIV positive individuals had an overall greater risk (OR 2.17), but that some studies showed a negative effect. There are case reports of HIV and *Taenia solium* (cysticercosis) co-infection (Chianura et al. 2006; Delobel et al. 2004), but little is known regarding the epidemiologic relationship between the two.

Independent of the number of people living in a household and sex and age, we observed a positive relationship between the species richness for the neglected helminths of interest and individual infection with HIV. This is quite striking, particularly given that a positive effect was also observed with raw (unadjusted) data, as well as in the null model, suggesting it is not only be an artefact of the variables included in the community model. A relationship between household-level infection with multiple (or indeed any) helminth infection has not, to our knowledge, been previously explored. There is substantial and increasing evidence that the three STH and schistosomiasis can increase susceptibility to HIV (described at length in the literature review) and it could be hypothesised that household helminth infection impacts upon within household HIV transmission dynamics. Much more work is needed to explore this association, but it could be hypothesised that strategies to deworm HIV positive individuals in addition to their families might contribute to reduced within-household HIV transmission.

Several limitations need to be considered, particularly in the light of this potentially important relationship with HIV. Whilst random effects have been incorporated to model a common species effect, they have not been incorporated at the household-level to account for between household heterogeneities in detection. As described in detail in the previous chapter, residual correlation amongst individuals in the same household may introduce bias into these models.

Our analysis has considered only the affect of the social and biophysical conditions on the probability of parasite occupancy, as well as household-level and individual effects on detection. The possible impacts of the pathogen interaction at the individual level has not been considered. It is quite possible that the presence of one species could influence the probability of detection of another parasite, by influencing susceptibility to infection, persistence or fecundity (Janovy 2002). Such an effect would violate the assumptions of the model, which requires species independence, and may lead to over-inflation in counts (Iknayan et al. 2014).

Key findings

1. Other than household size, there is little evidence that the social and environmental conditions considered here have a consistent effect on household-level infection with the neglected helminths of interest;
2. Despite this, strong evidence of spatial autocorrelation suggests there may be geographically heterogeneous social and environmental factors that influence household (and implicit in that, individual) risk of infection that have not been accounted for;
3. Household-level infection with multiple neglected helminths can be considered to be the norm rather than exception in this mixed farming community in western Kenya;
4. Socioeconomic position has little effect on the probability that a household is infected with the neglected helminths of interest, but individual infection, or intensity of infection, seems more likely in poorer households for several infections;
5. Helminth species richness (or household polyparasitism) was a positive predictor of individual HIV infection. Further work is needed (not least to improve the model predictions through better control for within household correlation), but this suggests HIV transmission occurs more readily in households with multiple helminth co-infections. Whilst we have done little here to explore causality, these findings provide some support to previous calls for integrated NTD and HIV control.

Chapter 8

Discussion

"*The natural locus of disease is the natural locus of life - the family.*"

Michel Foucault. The Birth of the Clinic, 1973.

Neglected tropical diseases are defined by their occurrence in poor, marginalised communities, but not all poor communities, nor all of the people within them, are likely to be at equal risk. Describing and explaining heterogeneities in risk is at the heart of modern epidemiology, and has enabled the discipline to inform public health interventions that seek to optimise the allocation of scarce resources (Woolhouse et al. 1997). As I have shown in Chapter 4 of this thesis, and as is widely observed in the literature, conditions that influence infectious disease outcomes in individuals often do so through effects at the household-level. People living in different households in the same community may have very different risk of infection with a particular agent, but this risk will tend to be more similar for those people in the same household. Multilevel studies allow quantification of the effect of these household (and other) level risk factors on individual infection, and may therefore help to explain between-household variation in individual risk, or why an infection may tend to cluster within particular households (Diez-Roux & Aiello 2005; Merlo et al. 2006). In Chapter 5, I used a multilevel approach to explore the effect of household-level predictors of infection for several endemic infectious diseases, and was able to show that an individual's risk of infection with hookworm, *Plasmodium falciparum*, or *Entamoeba histolytica/dispar* reduces as the socioeconomic position of the household in which they live increases, whilst for TB, the reverse was true. Hence, if we sampled 10 people in a randomly selected household, and 10 people in a richer household, we might expect (on average, and net of any other predictors) to identify more cases of hookworm, malaria or *Entamoeba* spp. in the poorer household, or that the (average) within-household prevalence would be higher (or lower in the case of TB). Therefore, for these three parasites, it would be reasonable to suggest that infection will be most common in the poorest households. We could not say, on the basis of these results, that poorer households are more likely to have *any* infection at all. This is perhaps a subtle distinction, and whilst the two states (of none and some versus different levels of some) are likely to be positively correlated, they can generally be considered to be

two distinct processes, and may therefore have heterogeneous relationships with covariates (Kery 2008; MacKenzie & Nichols 2004).

A similar distinction could be applied to individual-level infection with the STH or with *Schistosoma* species. Given that there is no replication within the host, studies of the transmission dynamics of these parasites have tended to focus on the intensity of infection rather than on presence or absence (Bundy & Medley 1992; Anderson & May 1992). It is also widely observed that infections tend to be highly aggregated, so that whilst most people have little or no infection, a small proportion may have very heavy infection burdens (Woolhouse et al. 1997; Guyatt et al. 1990). The most intense infections are generally associated with the greatest morbidity (Brooker et al. 2008; Pullan & Brooker 2008), and therefore work should certainly continue to focus on explaining individual variation in infection intensity. However identifying the factors that may influence whether or not an individual has *any* infection may also be useful. Such studies may help to better understand transmission within an endemic landscape, particularly when heterogeneous effects on infection and infection intensity are described (Chipeta et al. 2013). There are, of course, a large number of studies (including the one in Chapter 5) that have defined infection on the basis of detection or non-detection of helminth eggs in faeces, and used that binary outcome in a logistic regression to explore predictors of infection. However, given that the probability of detection of at least one egg in a faecal sample is an intensity related process, it could be expected that factors that influence the probability of detection are not necessarily the same as those that influence the presence or absence of infection.

In this thesis, I have considered infection at the household-level, or the dichotomy between presence or absence, to be an outcome of substantive interest. By examining infection risk at both the individual- and household-level, I have been able to provide additional insight into the landscape ecology and epidemiology of a number of infectious diseases within a single community than would have been possible by considering infection at either scale in isolation. Moreover, I have described the application of zero inflated binomial regression as a reasonably straightforward means with which to define infection status at the household level while accounting for the observation process in household members. This technique is very widely used in ecology for a similar purpose, but has been comparatively "neglected" in NTD research, a field in which I believe it has substantial utility.

The following sections provide a summary of the main findings in this work, and extend some of the ideas that were introduced in the discussion in each chapter.

8.1. The spatial distribution of household infection risk

In Chapter 4, I used the presence or absence of infection in at least one household member to define household-level infection. This allowed me to explore the spatial distribution of household infection risk, and to perform significance tests for spatial clustering of household-level infection. The aim of this work was predominantly descriptive and exploratory, and it provided a useful basis for further hypothesis testing as part of this thesis, as well as directing future work. In particular, the somewhat 'quick and dirty' kernel smoothing approach allowed me to compare smoothed risk surfaces for a large number of different pathogens, and to describe correlations between pathogen pairs. Of particular interest were the observed relationships between *Ascaris*, *Trichuris* and TB. As was discussed in Chapter 2 , an immunological basis for interaction between helminths and TB at the individual level has been widely described, but there is relatively little field-based evidence to support an effect at the population-level. Further work should seek to explore this interesting finding with a particular focus on control for potential confounders. A particular challenge in this analysis will be the possible effect of helminth infection on the performance of the gamma-IFN assay in the diagnosis of TB (as described in Chapter 5), which could negatively impact upon our ability to identify positive relationships at the individual level.

A potential consideration for the work described in Chapter 4, and one that follows on from the comments in the introduction to this chapter, is that whilst the outputs were interpreted as representing the spatial distribution of household risk of infection, or household prevalence, it is quite possible that the spatial heterogeneity observed is confounded by factors that influence within-household prevalence. Given that diagnostic tests were imperfect, the probability of detecting at least one positive person in an 'infected' household (or the 'household' sensitivity) can be expected to be a function of the within household prevalence (Christensen & Gardner 2000); as this increases, the likelihood of correctly identifying one person with infection also increases. The use of the hypergeometric sampling procedure, and resulting derivation of a distribution of risk surfaces, provided some account for household-level misclassification, but all significance testing (using the spatial scan statistic and the KDE relative risk procedure) was performed using the observed data only. Hence the underlying contextual or compositional effects that contribute to the observed "hot spots" in infection risk may well be acting on within-household prevalence (and therefore individual risk and the probability of detection) rather than household-level infection (and therefore household risk), or (perhaps more likely) a mixture of the two. This may be merely an

academic point, since identifying where the 'most' disease occurs may be the primary interest from a policy and disease control point of view. However, for future extensions of this work, and particularly where the aim is to truly represent the distribution of a species within a landscape, issues related to the probability of detection should be taken into account when modelling the distribution of household-level infection (Rota et al. 2011; Royle et al. 2012).

Perhaps the most interesting output from the spatial analysis in Chapter 4 was the indication of an area of elevated risk for multiple infections in the southern half of our study area. That this general effect was repeated using three separate approaches (i.e. the "probability of at least n" method, spatial scan statistic for household pathogen species richness, and examination of the division level residuals from a multilevel Poisson regression) makes this finding quite compelling. The analysis in Chapter 7 was intended to identify contextual or compositional effects that might operate on household risk for multiple pathogens, and which might therefore help to explain this effect. Unfortunately, other than household size, that analysis did not reveal a consistent 'group' effect of any of the hypothesised predictors on the helminths under study. Potential modifications to the analytical approach used are therefore worth considering. It may be useful, for example, to extend the hierarchical community models described in Chapter 7 to a wider range of parasite species (for example, by including the gastrointestinal protozoa). This could include the use of a stratified analysis of closely related or functional groups of pathogens (to avoid the problems of the smoothing of group level regression coefficients due to heterogeneous species effects that were described in Chapter 7).

A potentially more straightforward approach to explore the factors that influence species richness to that adopted in Chapter 7 would be to model the count of unique pathogens at the household level using Poisson regression, as we did for the multilevel model in Chapter 4 (and therefore seek to explain the apparent residual spatial autocorrelation observed in that chapter using covariates). Whilst this approach would be considerably simpler to implement than the hierarchical community model, it would not allow imperfect detection for each species to be accounted for. Given that diagnostic sensitivity for the gastrointestinal helminths is considered to be quite low (reviewed in Appendix 4.2), this could be expected to result in biased estimates of species richness, and potentially of covariate effects (Kery & Royle 2008; Popy et al. 2009). Moreover, individual species effects would not be separable. Hierarchical community models address both of these issues, whilst allowing inference on the same outcome, i.e. unique species counts (Kery & Royle 2008; Dorazio et al. 2011).

8.2. Relationships between helminths and HIV

Further work to try to explain the observed geographic gradient in multiple pathogen risk is perhaps made even more worthwhile in the light of the observed spatial overlap of an area with significantly elevated household pathogen counts (on the basis of the spatial scan statistic) and the HIV spatial "hot spot" of elevated (household) relative risk.

This finding, as well as that describing relationships between HIV and helminth infection at the individual level, and the apparent contextual (household level) effect of helminth species richness on individual risk of HIV infection, are potentially important outputs from this work.

In support of much of the published literature, the probability of detecting helminth infection (particularly hookworm and *T. trichiura*) on the basis of a single faecal sample was reduced in those individuals who are HIV positive (described in Chapter 5 and 6). It is not clear whether these individuals are less likely to be infected, or if they tend to have less intense infections, and therefore the probability of detection is reduced. This consideration provides a good example of the utility of considering individual infection, in addition to infection intensity, as an outcome of substantive interest. Zero inflated negative binomial models applied at the individual-level might help to shed some light on this question. These models would allow us to consider individual infection and intensity of infection as two part (linked) process, whilst allowing control for a range of potential confounders (and age and sex are likely to be important).

Despite the apparent negative relationship between HIV and helminth infection at the individual level, people living in households that have higher helminth species richness appear to be at greater risk of HIV infection. We have not explored this relationship in great depth, and much more work is needed to try to tease out potential confounders, as well as whether the magnitude of the observed effect may be due to a particularly strong relationship with a particular helminth species. The results from Chapter 6, where the number of people with HIV in a household (with adjustment for the number of individuals sampled) was an important predictor of detection for *Ascaris lumbricoides* and *Trichuris trichiura*, suggests these two STH may be particularly important in driving the observed relationship. There was also substantial overlap in the area of high relative risk for HIV and *T. trichiura*, and to lesser extent with that for *A. lumbricoides*, *S. mansoni* and *Taenia* spp..

Even in the presence of confounding or species-specific effects, the finding that household

helminth species richness is associated with individual HIV risk, on the basis of both the observed and adjusted parasite species counts (although there was only weak evidence to support the former), suggests at the very least that households in our study area may be dually burdened by HIV and multiple neglected helminth infections. This study therefore provides further support to the calls for greater integration of HIV and NTD control (Noblick et al. 2011; Hotez et al 2009a; Hotez et al. 2006a; Manne & Maciag 2011). The 'rapid impact package', widely recommended for the integrated control of NTDs (Molyneux et al. 2005; Hotez et al. 2006a), includes albendazole, praziquantel, and ivermectin, which would be expected to remove all of the helminths considered in Chapter 7 (the STH, *S. mansoni*, *T. solium* and *S. strercoralis*) (Molyneux et al. 2005). These treatments could be provided for around USD 0.50 per person per year (Noblick et al. 2011). There is increasing evidence (including that given here) for important relationships between HIV and helminth infection (described in detail in Chapter 2), and integrated control of both is likely to be cost effective, particularly in the light of recent commitments from the private sector for anthelmintic drug donation (WHO 2012c). Decisions on allocating resources should of course be based on a range of factors, but in our own study area, and in the situation that resources were limited (which is not necessarily the case in Kenya where nationwide deworming programmes are underway (Mwandawiro et al. 2013)), there would certainly be a strong argument for directing MDA in and around the area identified as being a hot spot for HIV (identified on the basis of the spatial scan statistic, or the KDE procedure: Chapter 4). This would provide coverage for areas at high risk for schistosomiasis, trichuriasis and to a lesser extent for ascariasis and taeniasis. It may also be reasonable to expect that *Schistosoma haematobium* would be prevalent in this swampy area towards Lake Victoria.

Recommendations have been given for the treatment of HIV infected individuals with anthelminthics to in order to delay disease progression in ART naive individuals (Gerns et al. 2012; Walson et al. 2009). The outputs from this study, and particularly those from the community model in Chapter 7, suggest that deworming the rest of the household of an individual with diagnosed HIV infection, and who lives in a helminth endemic area, could also have some benefit. This suggestion is given somewhat tentatively, and further work is still needed, particularly to provide better control for the expected correlated nature of observations within the same household (as described in Chapter 6 and 7).

8.3. Zero inflated binomial regression: a tool for surveillance?

It is well known that the diagnostic tools for many of the NTDs, most notably the STH and schistosomiasis, are inadequate, and therefore the development of new tests with better performance should be a priority (McCarthy et al. 2012). This would assist with the accuracy of observational based research, as well as for surveillance and monitoring. A major issue with the application of currently available tests is that whilst they perform reasonably well in individuals with intense infections, they do very badly in individuals with low intensity infection, who tend to shed eggs infrequently and in low concentrations (more detail is given in Appendix 4). A recent meta-analysis derived average estimates of diagnostic test performance for the STH (Nikolay et al. 2014) but, as the authors show, these estimates are likely to be highly variable between populations, and will depend in particular on the underlying intensity of infection in the population under study. Very often this is not known *a priori*, and in such cases estimates of prevalence adjusted on the basis of predictions of diagnostic test performance will (or should) be very imprecise. When repeat samples are collected from at least some people during a survey, a possible solution to this would be to use a zero inflated binomial regression approach. As described in the discussion in Chapter 6, this method would allow a population-level average of diagnostic test performance (albeit only for sensitivity) to be estimated *within* the population to which inference is being applied. As we did for household level prevalence in both Chapter 6 and 7, this estimate could then be used to derive the probability that any individual who tests negative for a particular condition is truly infected, with these results aggregated at the community level to derive an adjusted estimate of prevalence. Particularly powerful is that the probability of detection can be conditioned on the basis of covariates, and could therefore be allowed to vary from individual to individual depending on their age, sex and any other factor considered important. As we show in chapter 6, these models are reasonably straightforward to fit, with several dedicated, open source packages available.

The use of these models to derive an equivalent estimate of diagnostic sensitivity has been previously described in wildlife disease research (Schmidt et al. 2013; Lachish et al. 2011), and, I believe, they have great potential utility in disease surveillance activities for human (and animal) health.

8.4. Effects operating on helminth transmission in the domestic and public domains

A major aim of the work in this thesis was to describe the importance of within- and between-household transmission, particularly of the neglected helminths, and to try to

disentangle some of the factors that may influence risk of infection at these different levels. As already described, substantial household-level clustering was observed for several infectious agents, and particularly for each of the neglected helminth species under study (viz. *A. lumbricoides*, *T. trichiura*, hookworm, *Taenia* spp., and *S. mansoni*). The partitioning of variation in individual risk of infection at this level suggests factors operating at the household are important in explaining individual risk of infection, and therefore clustering within particular households. We focused primarily on household socioeconomic position (SEP) as a potential explanation for this clustering, and whilst associations were found between individual infection with hookworm (but not *A. lumbricoides* and *T. trichiura*) and SEP in Chapter 5, this predictor explained only small amounts of the variation in risk between households. Further work would therefore be valuable to explore alternative predictors of the observed between-household variation; genetic factors in addition to social and environmental effects are likely to be particularly important (Ellis & McManus 2009; Pullan et al. 2010b; Pullan et al. 2008; Walker et al. 2011). The presence of high levels of clustering at the household level strongly suggests within-household (domestic) transmission is important for the helminth species considered, a finding commonly reported in the literature (Walker et al. 2011; Pullan et al. 2010b; Brooker et al. 2006a; Forrester et al. 1988). It is perhaps worth noting, however, that household-level clustering could also occur if household level factors meant household members were repeatedly exposed to infection in the public environment (i.e. outside the household), and therefore repeatedly introduce infection. Separating the effect of household-level factors on individual infection risk inside and outside the household (or on transmission in the domestic and public domains) is likely to be difficult when infection outcomes are measured at the individual level only (and at a single point in time). This is likely to be particularly problematic if risk factors have heterogeneous effects in each environment.

In addition to within-household clustering, we show that between-household transmission of helminth (and other) infections is also likely to be important in the study area. Indeed, the presence of spatial clusters for each species of interest (as identified using the kernel smoothing approach for relative risk described in Chapter 4) suggests that there are factors that operate on rates of between-household transmission in several parts of the study area. By considering household-level infection as a substantive outcome, and assigning this with control for the observation process within households, we were able to quantitatively explore some of the factors that may influence this spatial clustering. Drawing on the results from Chapters 5 and 6, Table 8.1 provides a summary of the direction and magnitude of effect of covariates operating on outcomes measured at the individual level using multilevel logistic

regression (Chapter 5) and at the individual- and household-level using zero inflated binomial regression (Chapter 6) for the soil transmitted helminths. Comparison of these outputs indicates some potentially interesting heterogeneous effects of covariates at each level. For example, the NDVI, which has previously been shown to impact upon individual risk of infection of hookworm (Pullan et al. 2012), and which we also found was a risk factor at the individual level using multilevel logistic regression, appears to operate mainly on between household transmission, or infection in the public domain (since the covariate performs better as a predictor of household infection than individual infection within 'infected' households on the basis of the ZIB regression). Household SEP, by contrast, appears to have little effect on the probability that a household is infected (and therefore, presumably, on transmission in the public domain) but was a risk factor for individual infection in infected households. It would therefore appear that rich and poor households are equally likely to have any infection (or there is no evidence of a difference) but that most domestic transmission occurs in the poorest households. A similar finding was observed for the effect of household SEP on the probability of detection for *A. lumbricoides*, *S. mansoni*, and *S. stercoralis* (in addition to hookworm) in Chapter 7, strongly indicating the importance of household poverty on the domestic transmission of these neglected helminths. As described in Chapter 5, 6, and 7, the likely absence of sanitation in many of these households makes this finding unsurprising.

Examining infectious disease risk at multiple levels can also reveal covariate relationships that may be missed by focussing on the individual level only. For example, whilst local population density was not identified as a risk factor for individual infection with hookworm, there is weak evidence that it is associated with reduced risk of household level infection (Table 8.1). Hence, between household transmission, or infection in the public domain, may be greatest in those areas with lowest population density. Although this is perhaps the opposite to the expected effect, potential explanations for this finding could include greater exposure to infectious larvae of this soil transmitted parasite in fields in more rural, agricultural areas. Similarly, whilst there was some very weak evidence that maximum land surface temperature (LST) is negatively associated with *T. trichiura* infection at the individual level (on the basis of the multilevel logistic regression), there was strong evidence of a large positive effect at the household level. This suggests an important effect of LST on between household transmission, and that exposure outside the household may be greatest in those areas with highest temperature.

Despite these interesting findings, and as is shown in Table 8.1 and described in Chapter 6, we were able to identify relatively few significant effects of covariates operating at the household level on between household transmission. Moreover, we were unable to identify any shared effects of covariates on household level infection in Chapter 7 (other than household size). It may therefore be worth exploring the effects of a broader range of social and environmental factors on the probability of household infection, including higher level contextual effects such as indicators of community level poverty (potentially available from census data (e.g. Open Kenya (2014)), access to health services (and Kenya's health centres have been mapped (Noor et al. 2009; Open Kenya 2014)), and the degree of local infrastructural investment. Government based development funding to Kenya's 290 constituencies is currently managed through the Constituency Development Fund (CDF) (http://www.cdf.go.ke/). This is a publicly accountable scheme which, in theory, provides open access information on the annual allocation of all monies spent and their purpose per constituency. These data could potentially be explored and information on health, water and sanitation projects extracted and used to explore these higher level contextual effects.

It may be worth noting that there are differences in the size (but not direction) of effects of predictors at the individual level when comparing the multilevel logistic regression and ZIB models, most notably age for hookworm and *A. lumbricoides* (Table 8.1). The former method includes a random effect at the household level, whilst the latter does not, which may explain these differences. Alternatively, this may provide further evidence of the importance of different age groups on within household (domestic) transmission. In particular, the multilevel logistic regression demonstrated a negative relationship between age and *A. lumbricoides* risk, and therefore we would expect that children would be most likely to be infected. The same relationship was observed at the individual level in the ZIB model, but the effect was much more substantial. The youngest children tend to have the most intense infections for this parasite (Walker et al. 2011), and it has been suggested that they are the most common source of infection in household transmission (Killewo et al. 1991). The strong effect observed at the individual level in the ZIB model (with zero inflation on the basis of household infection) supports this, and we would expect that the probability of detection (i.e. infection) in infected households would be greatest in those that have the youngest children. Similarly, we repeat the findings from the literature by showing that risk of hookworm infection increases with age (albeit non-linearly) (Brooker et al. 2004a). Again, this effect was more substantial at the individual level in the ZIB model. In future work, considering household age composition as a predictor of both the probability of detection and household infection (occupancy) may therefore be useful to explore the

relative contribution of different age groups on within and between household transmission. As described in Chapter 5, future work will also explore the importance of including a household random effect in these ZIB models.

8.5 Implication of study findings for disease control in western Kenya

Dividing transmission events into household- and between-household domains has relevance for disease control activities, since different interventions may be appropriate in the different environments (Cairncross et al. 1996). As has already been described, other than somewhat intractable conditions such as NDVI, maximum LST, population density and the number of people in a household, we were able to identify relatively few factors that influence the between household transmission of infection for the helminth infections. These and other indicators could potentially be used to direct control activities toward high risk areas within endemic communities in an effort to reduce between-household transmission. However, the extent to which household-level control for endemic infections (and particularly the helminths) could ever be operationalised has been questioned (Brooker et al. 2006a). Moreover, and concerns about resistance notwithstanding, the fact that most treatments for the helminthic NTDs are very cheap means treating whole communities in order to reach high risk individuals or households may well be more cost effective than trying to seek them out for targeted interventions.

Policy makers should be aware that for several infectious agents (notably malaria, hookworm and *Entamoeba* spp.), the risk of individual infection is higher in poorer compared to comparatively richer families, even in a rural farming community that may appear uniformly poor. Moreover, the risk of within household transmission for hookworm, *A. lumbricoides S. mansoni*, and *S. stercoralis* appears to be highest in the poorest households. Hence, structural interventions that seek to reduce household poverty, such as the promotion of household latrine provision, could be expected to reduce the burden of helminthic infections within the poorest households.

Table 8.1. Comparison of outputs for outcomes measured at the individual level using multilevel logistic regression (MLR) and at the individual and household level using zero inflated binomial regression (ZIB). Dashes are shown where the variable was not included in the final model.

	Hookworm			***Ascaris lumbricoides***			***Trichuris trichiura***		
	MLR	**ZIB**		**MLR**	**ZIB**		**MLR**	**ZIB**	
	Individual	*Individual*	*Household*	*Individual*	*Individual*	*Household*	*Individual*	*Individual*	*Household*
Individual-level									
Age	1.27 (1.15 – 1.40)	1.82 (1.57 - 2.14)	-	0.84 (0.78 - 0.90)	0.59 (0.48 - 0.74)	-	0.92 (0.88 - 0.97)	0.79 (0.67 - 0.93)	-
Age^2	0.99 (0.98 – 1.00)	0.82 (0.74 - 0.91)	-	-	-	-	-	-	-
Sex	1.31 (1.06 – 1.61)	1.38 (1.12 - 1.70)	-	-	-	-	0.65 (0.47 - 0.91)	0.6 (0.44 - 0.84)	-
HIV	0.31 (0.17 – 0.56)	0.28 (0.16 - 0.49)	-	0.69 (0.24 - 1.96)	0.54 (0.22 - 1.36)	-	0.22 (0.1 - 0.52)	0.3 (0.14 - 0.63)	-
Household-level									
SEP	0.64 (0.55 – 0.76)	0.66 (0.58 - 0.75)	-	0.85 (0.60 - 1.21)	0.67 (0.54 - 0.84)	-	0.95 (0.76 - 1.20)	-	-
NDVI	1.22 (1.01 – 1.48)	-	1.62 (1.17 - 2.25)	-	1.25 (1.01 -1.54)	-	-	-	-
$NDVI^2$	0.83 (0.73 – 0.94)	-	-	-	-	-	-	-	-
Maximum LST	-	-		-	-	-	0.80 (0.62 - 1.02)	0.78 (0.66 - 0.92)	2.46 (1.48 - 4.10)
Pop. density	-	-	0.72 (0.52 - 1)	-	-	-	1.39 (1.12 - 1.72)	1.65 (1.23 - 2.20)	-
Pop. $density^2$		-						0.93 (0.86 - 1.00)	-
HIV count	-	-	-	1.52 (0.93 - 2.49)	1.22 (1.03 - 1.45)	-	1.53 (1.13 - 2.08)	-	-

8.6 Future directions

Throughout this discussion, I have highlighted several areas in which the methods used in this thesis could be extended in order to further explore potentially interesting questions. These could include:

1) The modelling of helminth infection at the individual level using infection intensity. Zero inflated negative binomial models would also allow an infection (prevalence) outcome to be modelled and would represent a very efficient use of our data;

2) The use of geostatistical models to explore the spatial co-occurrence and co-intensity with control for confounders and presentation of associated uncertainty. This could initially focus on those pathogens for which spatial relationships were described in this thesis, and could also include zero inflation to model both co-infection and co-intensity (*sensu* Soares Magalhães et al. (2011b))

3) Greater control for the correlated nature of outcomes in the same household through integration of random effects into both the single species and community site occupancy (ZIB) models.

4) Incorporation of a wider range of environmental and social effects operating at the household level in order to better explain both within and between household transmission of infection.

Appendices

Appendix to Chapter 4

A.4.1. Age infection profiles

Different age groups may have different prevalence profiles for different age groups due to differences in behaviour and exposure, as well as due to immunological differences. These considerations are important in relation to individual polyparasitism (although our main focus in this chapter was household polyparasitism).

The age pyramid for the study area is shown in Figure A.4.1

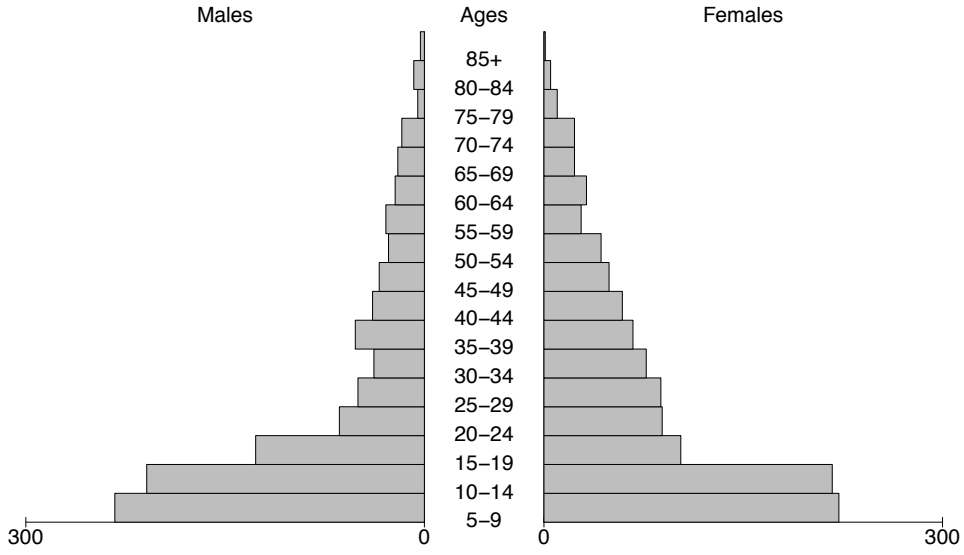


Figure A.4.1 . Age pyramid for sampled population

Adjusted prevalence for different age-groups (5 to 9 years, 10 to 14 years, 15 to 24, 25 to 39, 40+, which produces roughly equal group sizes) and male and female gender were derived for infections with an (unadjusted) individual prevalence >5% (i.e. more than around 100 cases).

Figures A.4.1 to 3 show the age profiles for the common infections.

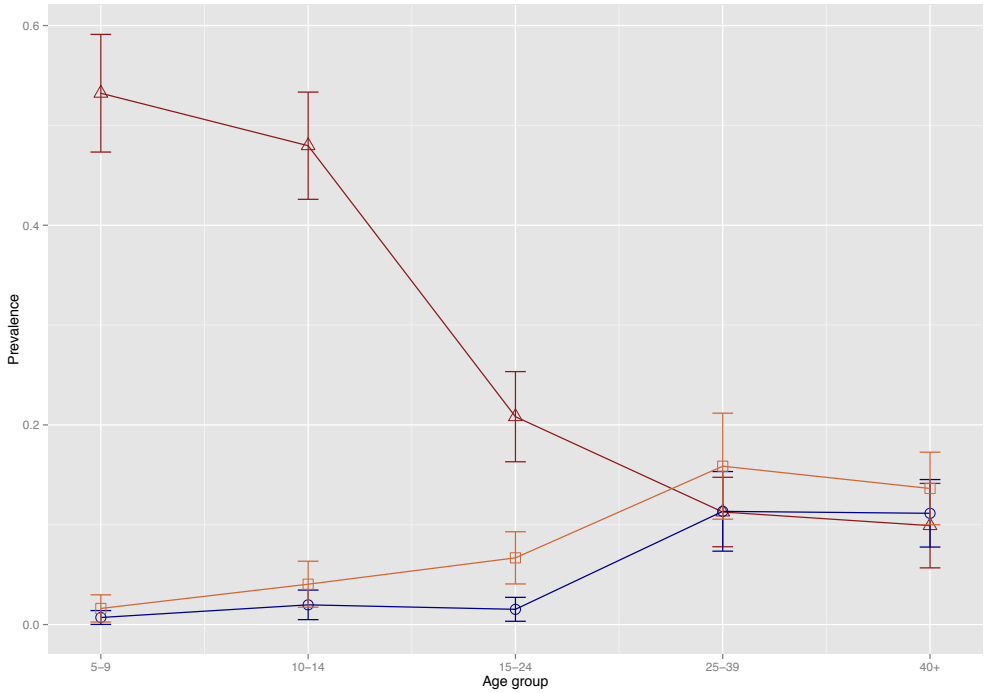


Figure A.4.1. Age-stratified prevalence of *Plasmodium falciparum* **(**Δ**),** HIV (O) and TB (□).Point estimates and standard errors (with error bars representing 95% CI) adjusted using design-based inference (see main text in Chapter 4)

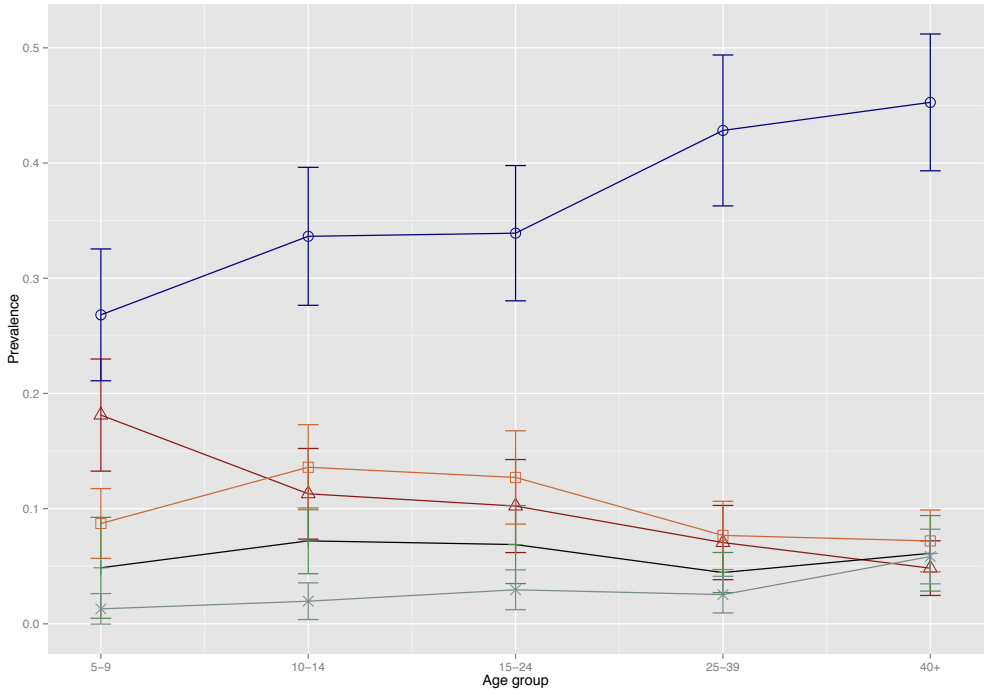


Figure A.4.2. Age-stratified prevalence of *Ascaris lumbricoides* **(**Δ), hookworm (O) and *Trichuris trichiura* (□), *Schistosoma mansoni*(×) and *Strongyloides stercoralis*(+). Point estimates and standard errors (with error bars representing 95% CI) adjusted using design-based inference (see main text in Chapter 4)

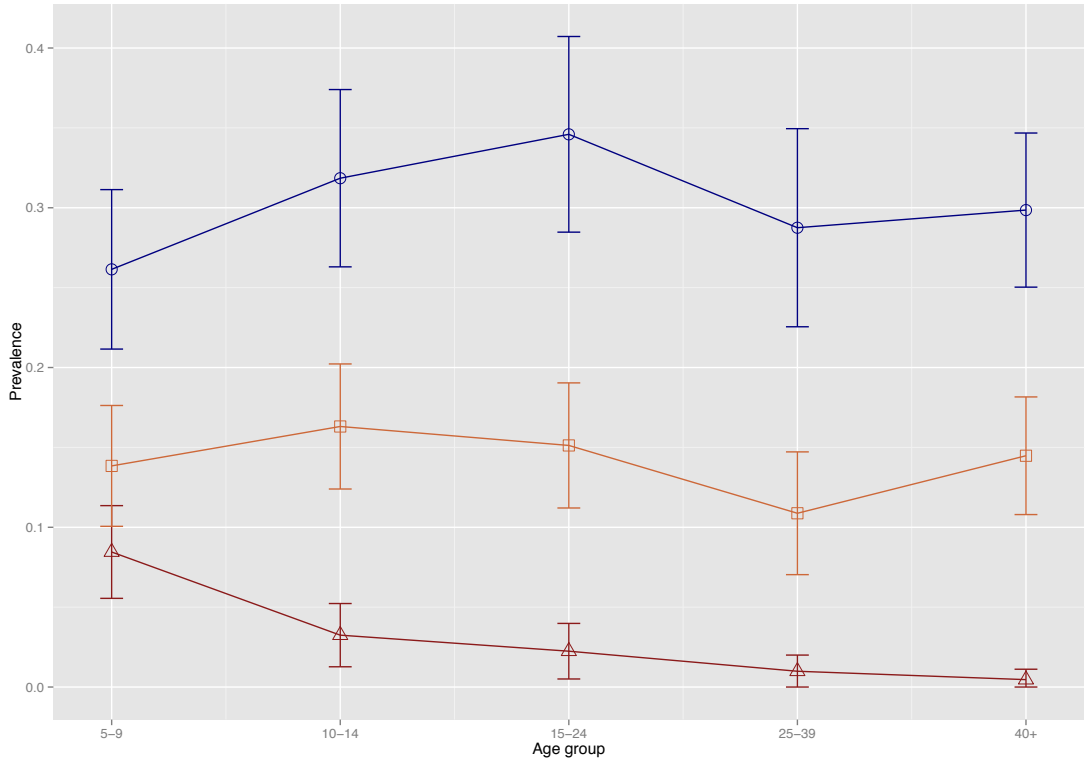


Figure A.4.3. Age-stratified prevalence of *Giardia* spp. (Δ), *Entamoeba histolytica/dispar* (O) and Iodamoeba butschlii (□). Point estimates and standard errors (with error bars representing 95% CI) adjusted using design-based inference (see main text in Chapter 4)

A.4.2. Literature review of diagnostic test performance

In the following section, we review the performance of the diagnostic tests used to inform individual and household prevalence estimates reported in Chapter 4. On the basis of this review, and in order to derive the Beta priors for sensitivity and specificity that were used in the hypergeometric group prevalence procedure, we estimate the modal (most likely) estimate of sensitivity and specificity for each test, as well as the value above we are 95% certain that the true value lies. These values are summarised in Table A.4.1.

Malaria

As with all microscopic techniques in which infectious agents are directly observed, microscopy for *Plasmodium* infection is subject to user error, and performance can be expected to vary according to the expertise of the technician performing the test (Ohrt et al. 2007; Wongsrichanalai et al. 2007). Moreover, diagnostic performance may vary in clinical versus community settings (Ochola et al. 2006), in low versus high transmission areas (Okell et al. 2009), and according to the number of

microscopic fields read (Trape 1985). Literature-based estimates of test performance are therefore variable and difficult to derive, particularly since expert microscopy is commonly used as the reference test, rather than as the test under study.

Our case definition for malaria infection was the observation of malaria parasites (including gametocytes) in either a thick or thin blood smear on microscopy, since this was the only test performed. Comparisons with PCR, which can detect parasites at much lower intensities (Roper et al. 1996), therefore reflect (unless appropriately adjusted) the sensitivity of microscopy to detect infection rather than parasitaemia above the theoretical limit for microscopy of 4-20 parasites/µl (Payne 1988): the prevalence on the basis of microscopy may be 50% that by PCR, particularly in low transmission settings (Okell et al. 2009). Test performance studies in which PCR is used as a gold standard tend to report low sensitivities for microscopy, of the order of 46% in Uganda (Batwala et al. 2010), 52% in Ethiopia (Santana-Morales et al. 2012) and 60% in Angola (Fançony et al. 2013). Sub-microscopic infections probably rarely result in clinical episodes (Rogier et al. 1996), although they may be important for, and of course reflect, parasite transmission (Roper et al. 1996). When the performance of expert microscopy was assessed using slides of known status, sensitivity was 95% for microscopists who had completed an intensive period of training in Kenya (Ohrt et al. 2007), but was lower (70.6%) in a similar group in Uganda (Namagembe et al. 2012). Kilian et al. (2000) found an agreement rate of 84.9% amongst 4 expert microscopists reading 464 positive samples from Uganda, with, as expected, increased levels of agreement in higher intensity infections.

Using a latent class 'no gold standard' analysis within a Bayesian framework, Ochola et al. (2006) compared the performance of microscopy with antigen detection tests (which have a detection limit of 100 parasites/µl) and reported posterior means for sensitivity that ranged between 78.6 and 94.2% for data from 18 field-based studies in malaria endemic countries. Specificity was always above 96.6%, and in some cases as high as 99.9%. Speybroeck et al. (2011) performed similar analyses of data from surveys in low malaria incidence populations in Peru, Vietnam and Cambodia, and reported sensitivities that ranged between 53 to 90%. In all cases, the lower bounds of the credibility intervals describing specificity were above 94%.

On the basis of this evidence, and considering that all malaria diagnostics as part of this study were performed by one of two 'expert' microscopists, we were 95% certain that the sensitivity for detection of malaria parasitaemia (with >20 parasites/µl) as part of this study was above 70%, with a modal value of 90%. False positive malaria diagnosis on the basis of microscopy can occur, through, for example, misidentification of stain precipitates or dirt and cell debris (Wongsrichanalai et al. 2007). Such errors will depend on the quality of slide preparation and reading, and are likely to be rare in expert microscopy (Speybroeck et al. 2011; Ochola et al. 2006; Ohrt et al. 2007). We were therefore 95% confident that the specificity for malaria detection was above 95%, with a modal value of 99%.

Soil transmitted helminths and *Schistosoma mansoni*

We used the results of a single Kato-Katz (Katz et al. 1972; Cox 1992) and formal-ether test (Allen & Ridley 1970) in parallel to determine infection status with hookworm, *Ascaris lumbricoides*, *Trichuris trichiura* (the soil-transmitted helminths (STH)) and *S. mansoni*.

Given that the microscopic diagnosis of STH and *S. mansoni* infection is based on observation of ova in faeces, the sensitivity of microscopy using either the Kato-Katz (KK) or formal-ether (FE) approaches will depend on factors that influence the intensity of infection within an individual, and the fecundity of resident parasites (de Vlas & Gryseels 1992). These factors include, age, sex, co-infection, and nutritional status. Hence, the sensitivity of microscopy can be expected to vary substantially at the individual level. Moreover, ova may be shed intermittently, particularly in light infections, hence sampling effort will also impact upon test performance (Utzinger et al. 2001; Booth et al. 2003; Krauth et al. 2012). These factors go some way to explaining the very wide range of literature-based estimates of the sensitivity of microscopy for helminth infections using either Kato Katz or formal-ether. For hookworm, for example, Levecke et al. (2011) report a sensitivity of 78.3% for a single Kato-Katz performed using samples collected from surveys of school children from Brazil, Vietnam, India and Tanzania (with variation of between 51 and 92.3% between study sites) whilst Booth et al. (2003) reported a sensitivity of just 8% in a sample of school children in Cote d'Ivore. Some of the observed differences are also methodological: Levecke et al. (2011) compared the results from a Kato-Katz with a McMasters assay (two non-independent tests with imperfect performance) on the same faecal sample in order to determine infection status (and therefore the sensitivity of either test), whilst Booth et al. (2003) et al used a (more appropriate) Bayesian latent class 'no gold standard' analysis on the results of multiple samples collected from the same individual over consecutive days. Operator error in the diagnosis of helminth infections can also be substantial: Bogoch et al. (2006) report diagnostic sensitivities of 71% (95% CI 49-87%) and 17% (95% CI 6-38%) for *S. mansoni* using Kato-Katz on the same set of samples in two reference (i.e. presumably gold standard) diagnostic laboratories in Europe.

Whilst the majority of test assessment studies are performed in school-age children (Glinz et al. 2010b; Knopp et al. 2008) who, along with pregnant woman, tend to be at highest risk of morbidity from these infections, some authors have explored the diagnostic performance of these tests when applied to samples from community-based surveys, and therefore a cross-section of ages. Tarafder et al. (2010) used a Bayesian latent class, no gold standard analysis to assess the performance of a single, double and triple Kato-Katz on samples from a survey in a farming community in the Philippines. They report a sensitivity of 65.2% (60 – 69.8) for a single Kato-Katz for hookworm, 91.4% (90.5-92.3) for *T. trichiura* and 96.9% (96.1-97.6) for *A. lumbricoides*. The specificity was 93.8% (92.4 – 95.4), 94.4% (93.2 – 95.5) and 96.1% (95.5 – 96.7), respectively. Bogoch et al. (2006) estimate an overall sensitivity of 44% for hookworm and 89% for *S. mansoni* on the basis of a single Kato-Katz

using samples from a population of all ages in Cote d'Ivoire. They estimate sensitivity using a mathematical model (developed by Marti & Koella (1993)) parameterised by the frequency of positive results from multiple stool samples. Knopp et al. (2008) used the same mathematical approach to derive a sensitivity of 45% for a single Kato-Katz for hookworm, 63.4% for *T. trichiura* and 84.2% for *A. lumbricoides* in school-aged children in Zanzibar.

Few studies have quantitatively explored the improvements in sensitivity that can be achieved through the use of formal ether and Kato-Katz in combination. Using the composite result of a number of tests, Utzinger et al. (2008) report an individual sensitivity of a duplicate Kato-Katz and a formal-ether for the diagnosis of hookworm on a single faecal sample of 68.4% and 38.2%, respectively, in school-aged children in Cote d'Ivoire. The prevalence when considering both tests was higher than the individual tests alone: 51% on the basis of the Kato-Katz and 28.4% on formal-ether, or 58.8% in combination. Also in Cote d'Ivoire, and using a Kato-Katz and formal-ether on 3578 stool samples from school-aged children, Raso et al. (2006a) found a hookworm prevalence of 30.9% and 26.1%, respectively, and 43.3% in combination. Goodman et al. (2007), using samples from infants in Zanzibar, report a sensitivity of a duplicate Kato-Katz (compared to the composite result of 5 tests) of 81.5%, 77.1%, 96.5% for hookworm, *A. lumbicoides* and *T. trichuris*, while the sensitivity on the basis of FE was 77.8%, 62.5%, 51.8%. In combination, the sensitivity was 88.9%, 83.3% and 100%.

The formal-ether, which involves concentrating steps, may be a better test than Kato-Katz for diagnosis of *S. mansoni* infection, particularly in the presence of low intensity infection (Kongs et al. 2001). (Ebrahim et al. 1997) report the prevalence of low intensity (<100 epg) *S. mansoni* infection from a community based sample in Egypt to be 27.8% on the basis of a single Kato-Katz and 39.2% on the basis of formal ether, and 52% with combined techniques (which included up to four Kato-Katz slides). They report (on the basis of composite KK and FE results) a sensitivity of 53.6% vs. 75.8%.

We expect considerable variation at the individual level for all of the soil transmitted helminths and *S. mansoni*, but felt 95% confident that, at least at the population level, the true sensitivity of a combined KK and FE for the diagnosis of hookworm was above 50%, and that the modal value was somewhere around 70%. Tests appear to be more accurate for both *T. trichuria* and *A. lumbricoides,* probably in large part because hookworm eggs are extremely fragile and may be easily broken or altered (Utzinger et al. 2008; Dacombe et al. 2007). We therefore felt 95% confident that the true value for sensitivity was above 60%, with a modal value of 80%. We used the same values for *S. mansoni*.

Specificity is often assumed to be 100% for these tests, and of all the studies described above, only Tarafder et al. (2010) actively estimate specificities. As is the case with malaria, specificity could be expected to be influenced by the experience and motivation of the technician performing the study. We were 95% that specificity was above 95%, with a modal value of 99%.

Strongyloides stercoralis

Unlike the STH and *S. mansoni*, the ova of *S. stercoralis* hatch within the small intestine of their human host, and microscopic diagnosis is based on the observation of larvae in faeces (Ericsson et al. 2001). Infections may be light, and larvae are often shed intermittently (Uparanukraw et al. 1999), therefore detection is difficult, particularly on the basis of a single sample (Mirdha 2009; Montes et al. 2010). Diagnosis as part of PAZ was made using a single formal-ether test, which improves the yield of larvae compared to techniques such as Kato-Katz (Ericsson et al. 2001), but is known to be less sensitive than alternative coprologic approaches, such as nutrient plate culture and the Baermann concentration method (Ericsson et al. 2001; Uparanukraw et al. 1999). Despite (or perhaps because of) this, there are surprisingly few studies that have attempted to quantify the performance of formal-ether in the diagnosis of *S. stercoralis* infection, a parasite that seems to have been thoroughly neglected in the literature and amongst the public health community (Krolewiecki et al. 2013).

Uparanukraw et al. (1999) report sensitivities for a single formal-ether that range between 13 and 55% for detection of *S. stercoralis*. On a limited sample of 33 positive results derived from 5 different tests, Marchi Blatt & Cantos (2003) found a sensitivity of 48.5% for the examination of a single faecal sample. In China, Steinmann et al. (2007) report a 11.5% prevalence using results from a Baermann concentration test or Koga agar plate test amongst a random sample of 180 people, but were unable to detect a single infection using formal-ether. Using samples from Japan, Brazil and Thailand, and defining the reference result on the basis of four different coprologic techniques, Sato et al. (1995) report a sensitivity of just 45% for a single formal-ether examination. This extremely poor performance of a commonly used test (in surveys of gastrointestinal infection) will inevitably result in substantial underestimation of the prevalence of *S. stercoralis* (Bisoffi et al. 2013).

In the light of this (limited) evidence, we were 95% confident that the sensitivity of the formal ether for *S. stercoralis* infection was less than 50%, with a modal value of 40%. Few studies report specificity, but we expect this to be high and were 95% confident the value was above 95%, with a modal value of 99%.

Protozoal infections

As with helminth infection, excretion of trophozoites, cysts or oocysts from gastro-intestinal protozoal infection can be sporadic, making microscopic detection on the basis of 'O and P' (ova and parasite) examination challenging (Fotedar et al. 2007), particularly with a single faecal sample (Cartwright 1999). Moreover, a large number of human protozoal infections are commensal, and of no (known) clinical importance. Such infections can complicate diagnosis and management (Lawson et al. 2004). *Entamoeba histoytica,* for example, said to be the second leading cause of death due to parasitic disease world-wide (Stanley 2003), is morphologically (and therefore microscopically) indistinguishable from *E. dispar*, a non-pathogenic species (Utzinger et al. 1999). Studies in Cote d'Ivoire found that the vast majority *Entamoeba* infections among school children were with the

apathogenic *E. dispar* (Heckendorn et al. 2002). A similar finding was reported in Tanzania (Nesbitt et al. 2004), and indeed, globally, the prevalence of *E. dispar* infection has been estimated to be 10 times that of *E. histolytica* (Huston & Petri 1999). Infection with *Iodamoeba butschli* is also generally thought to be asymptomatic. However, such non-pathogenic gastrointestinal protozoa share the same transmission routes as the pathogenic species, and may therefore be useful as indicator species for shared risk factors (Giacometti et al. 1997; Abdolrasouli et al. 2009).

Cartwright (1999) examined the performance of a single 'O and P' for a range of protozoal parasites from a group of migrants to the USA, and found a pooled sensitivity of 75.9% using the composite results from examinations on at least three samples. This sensitivity increased to 92% when two samples were examined. Branda et al. (2006), also in the USA, report a similar sensitivity of 72% for all protozoa, when compared to the 'gold standard' of 3 microscopic examinations, whilst Suputtamongkol et al. (2006), in Thailand, report a sensitivity of 76.6%. Microscopy on a single stool specimen was reported to have a sensitivity of 85.7% for the detection of both *G. lambia* and *C. parvum* (with 95% CI of 62.6 – 96.2 and 56.2 – 97.5, respectively) and a specificity of 100% (with 95% CI of 96.3 – 100 and 96.5 - 100, respectively) in Saudi Arabia, (Elsafi et al. 2013); a composite reference test (including PCR and rapid immunoassay) was used as gold standard. In Australia, the sensitivity for single sample microscopy (again, compared to composite tests) of 73% for Giardia and 75% for *Entaemoeba* complex (Stark et al. 2014). Marti & Koella (1993), by contrast, report a sensitivity of 61% for *E. histolytica* [and *E. dispar*], and 75% for Giardia.

On the basis of this evidence, we were 95% sure that the sensitivity for all protozoa under study was greater than 70%, with a modal value of 75%. Specificity is often assumed to be 100% for the detection of ova, although species with similar morphologies can be confused (Stark et al. 2014), and immunological cells can potentially be mistaken as cysts and trophozoites (Fotedar et al. 2007). Clearly, and as with all coprologic techniques that rely on visual identificiation of parasite species, the experience and skill of the operator will have a major bearing on the specificity of microscopy. We were 95% confident that the specificity was above 95%, with a modal value of 99%.

Brucella* spp. and *Coxiella burnetii

For combined IgM and IgG immuno-chromatographic tests (as used in this study), the sensitivity has been estimated as 96% (88% – 99%) and specificity 99% (97% - 100%) by the original manufacturers of the assay (Smits et al. 2003). The sensitivity was found to be 100% (92.8 – 100) on a smaller number of culture confirmed samples in Kazakhstan (Mizanbayeva et al. 2009). Irmak et al. (2004) report that the level of agreement with a serum agglutination test with a titre of $\geq$1:160, often used as a 'gold standard' (although far from perfect) serologic test, was 92% whilst it was 100% for culture confirmed cases. They report a specificity of 95.8%. These estimates are based on individuals presenting with clinical disease: no validation studies have been performed in a community-based setting in which brucellosis is endemic.

There is limited data on the performance of the *Coxiella burnetii* Phase 2 ELISA, but it was estimated to have a sensitivity of 84.8% and specificity of just 63% to predict exposure in Togo (Dean et al. 2013), a sensitivity of 71% and specificity of 96% using samples from Australia (Field et al. 2002) and sensitivity of 97.7% and specificity of 100% using sera from Australia, Bulgaria and Slovakia (Slaba et al. 2005). In Spain, the sensitivity was 89%, and specificity 97% (Sanz et al. 2006). In all cases, the IFA was used as a gold standard. Only the Spanish study employed the same commercially available kit used in PAZ. The manufacturers of that kit report a sensitivity of 93.4% and a specificity of 98.5%.

For both *Brucella* and *Coxiella*, and as is the case for most antibody-based tests for pathogens that can be cleared by the immune system, a positive result only reflects exposure at some point. Currently there is little understanding of how far in the past that exposure could have occurred for infection with either species.

We were 95% confident that the sensitivity and specificity of LFAs for identification of antibodies to *Brucella* spp. were above 90%, with modal values of 96% and 99%, respectively (after Smits et al. (2003)). The ELISA for Q-fever seems to perform less well, although results are very variable, but we were 95% confident that the sensitivity was above 70%, with a modal value of 93% (after the manufacturers), and the specificity was above 60% with a modal value of 99% (after the manufacturers).

Tuberculosis

We used an interferon-gamma-release assay (IGRA) for the assessment of TB infection. These tests identify both latent and active infection, but are unable to distinguish between the two (Pai et al. 2004). In the absence of a gold standard for identifying latent infections, most literature based assessments of the performance of IGRA are made in comparison with active cases (Menzies et al. 2007; Pai et al. 2008). Such estimates, whilst all that are available for cross-sectional designs, may misrepresent the performance of these tests in latent infections which, by definition, exist in individuals without active infection (Kunst & Khan 2007). The estimation of specificity would be expected to be substantially reduced when the outcome of interest is the clinical state of 'active infection', rather than microbial infection (Metcalfe et al. 2011). A number of studies have shown variability in within-subject IGRA status over time (Ringshausen et al. 2012), although it is unclear if this represents true clearance, an underlying dynamism in the immunologic process, or poor test reproducibility (Zwerling et al. 2013).

IGRAs are used increasingly commonly in a clinical setting with the specific intention of identifying latent infection (e.g. prior to the introduction of immunosuppressive therapy (Ponce de Leon et al. 2008) or in immigrants from high incidence areas). However, the tests have been used relatively infrequently in cross-sectional surveys in a research setting, particularly in sub-Saharan Africa (Jensen et al. 2013; Legesse et al. 2011). Where they have been used, they have shown expected

epidemiological relationships, such as increased frequencies of IGRA positivity in individuals in contact with sputum positive TB cases (Jensen et al. 2013). Currently available IGRAs use *Mycobacterium tuberculosis* specific antigens, the genes for which are not present in the *M. bovis* BCG (vaccine) strain, or in other non-tuberculous mycobacteria (Menzies et al. 2007) We have been unable to find data to describe the accuracy of these tests to detect zoonotic *M. bovis* in people, although they have been used for that purpose in research activities (Torres-Gonzalez et al. 2013), and have been shown to detect individuals with active TB attributable to *M. bovis* in clinical settings (Singh et al. 2011).

There are now a very large number of systematic reviews and meta-analyses that have assessed the performance of IGRA's in a range of populations, and with a range of inclusion criteria. Of particular relevance for our own application, several of these included only those studies in which low risk controls were selected from low incidence populations, and in which specificity represents the test's ability to correctly identify uninfected individuals as test negative. They report a consistently high pooled specificity for the Quantiferon-Gold-in-tube test (QFT-IT, as used in the PAZ study) of 99% (95% CI 98 – 100%) (Diel et al. 2010) and 97.7% (95% CI 96 – 99%) (Menzies et al. 2007). Pai et al. (2008) report a slightly reduced specificity in vaccinated (96% (95% CI 94 – 98) compared to non-vaccinated populations (99% (95% CI 98 – 100). Sester et al. (2011), by contrast, included studies in which specificity was derived using individuals who were suspected to have active TB, but were diagnosed with a different disease, and report a lower specificity of 79% (95% CI 75 – 82).

The majority of systematic reviews explore the sensitivity of IGRA's using literature in which the outcome of interest was active TB, with various case definitions for 'active'. Metcalfe et al. (2011) focus on those studies performed in middle and low-income studies, and include data from Tanzania, Uganda, Zambia and South Africa. They report a pooled sensitivity of 60% (95% CI 45 – 92) from 5 studies of HIV infected individuals, and 84% (95% CI 78 – 91) in 9 studies of HIV un-infected persons. On the basis of a small number of studies, Chen et al. (2011) report slightly lower sensitivity in low/middle (77% (95% CI 70.1 – 82%) compared to high income countries (84.6% (95% CI 54.6 – 98.1) in HIV infected individuals. In children, pooled sensitivity was also lower in low and middle-income countries (70%, 95% 45 – 94) compared to high income countries (81% 95% 73 to 90). Although these differences are not statistically significant, the higher prevalence of helminth infection and malnutrition, both of which can have an immuno-modulatory effect, in low and middle income countries, could go some way to explaining these differences (Thomas et al. 2010). In children and adults from a range of populations (but predominantly in high income countries), sensitivity for predicting active TB cases is always suboptimal, with pooled estimates ranging between 76% to 89% (Dai et al. 2012; Menzies et al. 2007; Diel et al. 2010; Pai et al. 2008). Menzies et al (2007) also explore the performance of IGRAs using 3 studies that used a gradient of exposure as an indicator of the likelihood of latent TB infection (which perhaps best reflects our use for this test), and although it is difficult to ensure equivalence in terms of 'exposure' in each case, the prevalence of IGRA positivity was always highest in the most exposed groups.

We were 95% confident that the sensitivity of the test would be above 60% when applied to this cross-section of individuals with expected high levels of helminth infection and moderate HIV prevalence. We used a conservative modal value of 70% in order to predict sensitivity to predict all (i.e. latent and active) infection. For the specificity, we were 95% confident that the true value was above 95%, and chose a modal value of 98% as being the average of values derived from Pai et al (2008), Menzies et al. (2007) and Diel et al. (2010).

HIV

Rapid, highly accurate tests for HIV infection are widely available, and these properties are fundamental to control efforts. The test used in the PAZ study has been reported to have a sensitivity of 100% and a specificity of 99.4% in India (Vijayakumar et al. 2005), where it was able to detect antibodies in individuals to HIV-1 subtype A and C, the dominant subtypes circulating in Kenya (Khamadi et al. 2005; Khamadi et al. 2008).

We were 95% confident that the true sensitivity was above 99% with a modal value of 100%, and that the specificity was above 98%, with a modal value of 99.4%.

***Taenia* spp.**

Microscopic examination of stools for the presence of eggs due to infection with either *Taenia saginata* (the beef tapeworm) or *T. solium* (the pork tapeworm) has high specificity, but eggs are shed intermittently and the approach has a very low sensitivity (Flisser 2006; Praet et al. 2013). The copro-antigen ELISA, which aims to detect antigens to the *Taenia* genus (i.e. is not species specific) (Allan et al. 1990), was developed as a more sensitive alternative, but its performance has been explored only rarely in the literature. Allan et al. (1996) were able to identify *Taenia* segements (following treatment) or eggs in 70% of 79 copro-antigen positive people in Guatemala. Praet et al 2013 recently used a Bayesian non-gold standard approach on samples from two Zambian communities, and report a sensitivity of 84.5% (95% CI 62 – 98) and specificity of 92% (95% CI 90 – 94).

In the absence of additional information of sufficient quality, we used the posterior means and lower bound of the credibility intervals from the study of Praet et al 2013 to estimate the Beta distribution for both sensitivity and specificity.

Table A.4.1. Estimates of performance for the diagnostic tests used in the PAZ study on the basis of literature review

Infection	Diagnostic test	Modal value[1] (Sensitivity)	95% certainty[2] (Sensitivity)	Modal value[1] (Specificity)	95% certainty[2] (Specificity)	Prior Sensitivity (Beta(a,b))	Prior Specificity (Beta(a,b))
Cryptosporidium spp.	Ziehl-Neelsen staining	75%	>70%	99%	>95 %	Beta(174.5, 58.8)	Beta(88.3, 1.9)
Iodamoeba butschlii	Formal ether slide preparation	75%	>70%	99%	>95 %	Beta(174.5, 58.8)	Beta(88.3, 1.9)
Entamoeba histolytica/dispar	Formal ether	75%	>70%	99%	>95 %	Beta(174.5, 58.8)	Beta(88.3, 1.9)
Giardia spp.	Formal ether	75%	>70%	99%	>95 %	Beta(174.5, 58.8)	Beta(88.3, 1.9)
Strongyloides stercoralis	Formal ether	40%	<50 %	99%	>95 %	Beta(28.3, 42)	Beta(88.3, 1.9)
Schistosoma mansoni	Formal ether and Kato-Katz	80%	>60 %	99%	>95 %	Beta(14.8, 4.5)	Beta(88.3, 1.9)
Trichuris trichiura	Formal ether and Kato-Katz	80%	>60 %	99%	>95 %	Beta(14.8, 4.5)	Beta(88.3, 1.9)
Ascaris lumbricoides	Formal ether and Kato-Katz	80%	>60 %	99%	>95 %	Beta(14.8, 4.5)	Beta(88.3, 1.9)
Hookworm	Formal ether and Kato-Katz	70%	>50%	99%	>95 %	Beta(13.3, 6.3)	Beta(88.3, 1.9)
Taenia spp.	Copro-ELISA	85%	>62%	92%	>90%	Beta(12.1, 3.0)	Beta(576.6, 51.1)
Plasmodium falciparum	Thick and thin blood smear	90%	>70%	99%	>95 %	Beta(15.0, 2.6)	Beta(88.3, 1.9)
Brucella spp.	Rapid immunochromatographic assay (RIA)	96%	>90%	99%	>90%	Beta(71.1, 3.9)	Beta(34.2,1.3)
Coxiella burnetii	Phase 2 (IgG) ELISA	93%	>70%	99%	>60%	Beta(12.2, 1.8)	Beta(6.03, 1.1)
Mycobacterium spp.	Interferon gamma release assay	70%	>60%	98%	>90%	Beta(47.5,20.9)	Beta(42.1, 1.8)
HIV	RIA	100%	>99%	99.40%	>99%	Beta(298.1, 1)	Beta(1756.4, 11.6)

[1] The BetaBuster software requires as an input the most likely (modal) estimate for sensitivity and specificity, and [2] the value at which the user is 95% certain the true value is above or below.

A.4.3. R code used to create risk surfaces

The following R code was used to derive the distribution of spatial risk estimates for each infection (individually) using the package *sparr* when household infection is defined by a probability (in this case, derived using the BDFree package, outside of R).

```
################################################################
# kernel density estimation - simulated risk estimates e.g. HIV
################################################################

rm(list=ls())

library(maptools)
library(sparr)
library(raster)

paz = readShapeSpatial("~/outline_Project.shp")
# Shapefile of study area
paz = as.owin(paz)
hs_proj = read.csv("~/HS_project.csv", header = TRUE)
# Household co-ordinates (UTM)
fixed_bw = 5000 # Bandwidth for the kernel smoothing...
n = 3000 # number of iterations in simulation...should be big number

INF <- read.csv("~/HIV_pr.csv") # Household probability of infection

# make empty matrix for simulated infection status data
INF[,as.character(1:n)] <- NA
INF_m = as.matrix(INF[,c(1,5,6:(n+5))])

# do n bernouilli trials
for(i in 3:ncol(INF_m)){
  for(k in 1:nrow(INF_m)){
    INF_m[k,i] = rbinom(1,1,INF_m[k,2])
  }
}

# double check not massive difference between probability and
simulated
max(INF_m[,2]-(rowSums(INF_m[,3:(n+2)])/n))
min(INF_m[,2]-(rowSums(INF_m[,3:(n+2)])/n))

# denominator for kernel density (density all households)
INF_sim = merge(hs_proj, INF_m, by="HS_ID")
p_data = INF_sim[,c(2,3)]
p_kde <- bivariate.density(data = p_data, res = 200, adaptive =
FALSE, edgeCorrect = TRUE, WIN = paz, pilotH = fixed_bw,
                           intensity = TRUE, use.ppp.methods = TRUE)

# get number of rows for raster
y = raster(as.im(p_kde))
z = data.frame(rasterToPoints(y))
X = matrix(NA, nrow = nrow(z), ncol = n)

# simulation: ratio of each case column
for(i in 1:n){
  data = INF_sim[,c(2,3,i+4)]
  kde <- bivariate.density(data = data, ID = 1, res = 200, adaptive
= FALSE, edgeCorrect = TRUE, WIN = paz, pilotH = fixed_bw, intensity
```

```
= TRUE,  use.ppp.methods = TRUE)
  risk <- risk(f = kde, g = p_kde, plotit = FALSE, log = FALSE)
  x <- as.im(risk)
  ras = raster(x)
  p <- data.frame(rasterToPoints(ras))

  X[,i] = p[,3]
}

out = as.data.frame(X)
out$x = p$x
out$y = p$y

# get summary stats
out$avg <- rowMeans(out[,c(1:n)])
out$five <- apply(out[,c(1:n)], 1, quantile, c(0.05))
out$ninefive <- apply(out[,c(1:n)], 1, quantile, c(0.95))

av_xyz = out[,c((n+1):(n+3))]
five_xyz = out[,c((n+1),(n+2),(n+4))]
ninefive_xyz = out[,c((n+1),(n+2),(n+5))]

HIV_average.ras = rasterFromXYZ(av_xyz)
HIV_five.ras = rasterFromXYZ(five_xyz)
HIV_ninefive.ras = rasterFromXYZ(ninefive_xyz)

# plot it
brk = seq(from = 0, to = 1, by = 0.1)
plot(HIV_average.ras, col = heat.colors(11)[11:1], breaks= brk)
plot(HIV_five.ras, col = heat.colors(11)[11:1], breaks= brk)
plot(HIV_ninefive.ras, col = heat.colors(11)[11:1], breaks= brk)
```

The following R code uses the average estimate of risk for all infections of interest to derive a surface describing the probability of at least *n* infections over the surface of the study area.

The estimation is based on risk estimates for each cell of the rasters describing average risk of infection for each pathogen. These were first converted to points to facilitate combination.

```
all_human = data.frame(h_ma.p[,c(1:3)], h_as.p[,c(3)],
h_ho.p[,c(3)],
                       h_tr.p[,c(3)], h_sm.p[,c(3)], h_cr.p[,c(3)],
                       h_gi.p[,c(3)], h_eh.p[,c(3)], h_st.p[,c(3)],
                       h_hi.p[,c(3)], h_br.p[,c(3)], h_qf.p[,c(3)],
                       h_ta.p[,c(3)], h_tb.p[,c(3)])

colnames(all_human) <- c("x","y","Malaria","Ascaris","Hookworm",
                         "Trichuris", "S.mansoni", "Crypto.",
                         "Giardia","E.hist.",
                         "Strongy.", "HIV","Brucella",
                         "Qfever","Tapeworm","TB")

# derive probabilities…
# when inf is at least 2 less than n.
inf=2 #(to 12)
```

```
for(i in 1:nrow(all_human))
  {
N = as.numeric(all_human[i,3:16])
m = inf
A  = combn(N,m)
B = apply(A,2,function(S) setdiff(N, S))
y = apply(A,2,prod)*apply(1-B,2,prod)
X[i,inf] = sum(y)
}

# when inf is 1 less than n
inf = 13
for(i in 1:nrow(all_human))
{
  N = as.numeric(all_human[i,3:16])
  m = inf
  A  = combn(N,m)
  B = apply(A,2,function(S) setdiff(N, S))

  y = apply(A,2,prod)*(1-B)
  X[i,inf] = sum(y)
}

# when inf = n
inf = 14
for(i in 1:nrow(all_human))
{
  N = as.numeric(all_human[i,3:16])
  m = inf
  A  = combn(N,m)

  y = apply(A,2,prod)
  X[i,inf] = sum(y)
}

all_human$at_least_one = X[,1] + X[,2] + X[,3] + X[,4] + X[,5] +
X[,6] + X[,7] + X[,8] + X[,9] + X[,10] + X[,11] + X[,12] + X[,13] +
X[,14]
all_human$at_least_two = X[,2] + X[,3] + X[,4] + X[,5] + X[,6] +
X[,7] + X[,8] + X[,9] + X[,10] + X[,11] + X[,12] + X[,13] + X[,14]
……
all_human$at_least_thirteen =  X[,13] + X[,14]
all_human$at_least_fourteen =  X[,14]


# plot them
one = rasterFromXYZ(all_human[,c(1,2,17)])
…..
twelve = rasterFromXYZ(all_human[,c(1,2,2
```

Appendix to Chapter 5

A5.1 The spatial distribution of use of piped water in the study area.

We observed a lack of an expected relationship between urban distance and use of piped water and therefore explored the potential spatial distribution of household use of piped water using the spatial scan statistic in SatScan with a Bernoulli model (Kulldorff 1997). The output is shown in Figure A5.1 and reveals a cluster of piped water use in the south east of the study area around Lake Victoria. Within this cluster there were 25 households that use piped water, whilst there were 12 outside it (Relative risk = 5.65, p = 0.0020).

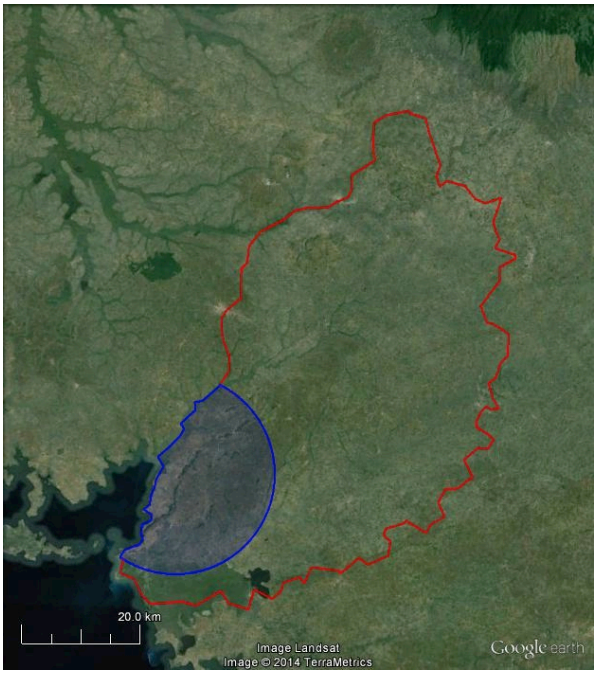


Figure A5.1. Cluster of piped water use.

A5.2 Scores and contributions of variables to the MFA.

The scores (i.e. principal co-ordinates) and the relative contribution (i.e. the quotient between the inertia of a point's projection and the inertia of all points on a particular axis, in this case the first) assigned to each variable in the MFA is shown in Table A5.1. The average household SEP value was 0, and therefore variables which have a negative score can be considered to make a negative contribution to SEP, whilst those with a positive score have a positive effect (i.e. a household is better off if it possesses them). In the case of material wealth, every asset owed has a positive effect on wealth, whilst not owning the asset makes a negative contribution. Similarly, not having access to

household services tends to result in a negative score, whilst access indicates higher SEP. Using spring water was considered to have a negative effect on SEP, and that those who do not were therefore 'better off'.

Table A5.1. Variable scores and relative contributions to the final MFA

Material wealth	**Scores**	**Contrib.**	**Access to services**	**Scores**	**Contrib.**
Cement floor	1.43	2.77	Power source	2.2	7.88
TV	1.58	2.67	Closed latrine	1.28	5.4
Brick/cement walls	1.52	2.31	No latrine	-1.06	5.06
No sofa	-1.01	2.25	No water treatment	-0.53	1.73
No cupboard	-0.89	2.08	No bed net	-1.15	1.57
No mobile phone	-1.26	2.04	Water treatment	0.37	1.21
Mobile phone charger	1.12	1.98	No power source	-0.27	0.96
Cupboard	0.79	1.86	Use piped water	0.81	0.87
Thatched roof	-1.01	1.78	Use borehole water	0.25	0.41
Sofa	0.65	1.45	Does not use borehole water	-0.2	0.34
Clock	0.82	1.44	Use well water	0.3	0.23
No bicycle	-0.94	1.42	Use spring water	-0.16	0.18
No radio	-1.26	1.42	Bed net	0.1	0.14
No torch	-0.74	1.22	Does not use spring water	0.13	0.14
Sewing machine	1.38	1.21	Does not use piped water	-0.08	0.08
Watch	0.91	1.03	Does not use well water	-0.06	0.04
Mud floor	-0.46	0.89	Partially closed latrine	0.08	0.04
No clock	-0.51	0.89	**Household resources**	**Scores**	**Contrib.**
Torch	0.49	0.81	Adults with sec. school ed.	1.04	4.92
Metal roof	0.46	0.8	1-2 adults in household	-0.95	4.33
No mobile phone charger	-0.44	0.79	>5 adults in household	1.06	2.54
Motorbike	1.48	0.65	No adults with sec. school ed.	-0.51	2.42
No TV	-0.37	0.63	External source of income	0.58	2.27
Mobile phone	0.38	0.61	No external source of income	-0.55	2.13
No bed	-1.13	0.56	4-5 adults in household	0.46	0.87
Bicycle	0.38	0.56	Adults<Children	-0.41	0.85
Mud walls	-0.33	0.5	Adults> Children	0.25	0.52
No watch	-0.26	0.29	Female household head	-0.19	0.2
Radio	0.24	0.27	Male household head	0.14	0.15
No sewing machine	-0.18	0.15	Household established <5yrs	-0.2	0.06
Bed	0.1	0.05	3 adults in household	0.1	0.03
No motorbike	-0.08	0.04	Household established >5yrs	0.03	0.01
Total livestock value	**Scores**	**Contrib.**			
TLV	0.55	15.02			

A5.3 Comparing outputs from multiple factor analysis (MFA) and the analytic hierarchy process for the estimation of socio-economic position.

Introduction

There is an element of the 'black box' to multivariate data reduction techniques, including multiple factor analysis (MFA). Moreover, some published SEP indices that used multivariate data reduction techniques report unexpected outputs, including counter-intuitive scores applied to variables (Houweling et al. 2003).

The biological 'realism' of the derived index can, of course, be assessed by examining the weighting assigned to each variable (as shown in Table A5.1) and assessing the extent to which it can be considered to agree with one's own subjective ranking (i.e. whether a score should be positive or negative and the magnitude of the score). Clearly, such a post-hoc assessment cannot be considered unbiased, and there may be a tendency to fit ones own perception of variable importance to the scores obtained. In order to ensure biological realism existed in the derived index, whilst minimising this bias, we compared the output from the objective MFA procedure with that derived from a subjective procedure in which variable scores were assigned on the basis of their assumed importance in defining socioeconomic position (SEP) in the community under study.

There are several ways in which SEP could be assessed subjectively, including participatory approaches to rank (and therefore derive scores for) assets, qualities or conditions representing SEP by members of the community of interest (Hargreaves et al. 2007), or through the use of targeted questions that ask participants to estimate their own SEP (Howe et al. 2011). In the absence of such data, we made use of the analytic hierarchy process (AHP), which has been used in a very wide range of disciplines and provide a means with which to make complex decisions from a set of competing elements (Saaty 1980). In our application of this process, these elements were the range of possible indicators of SEP (asset ownership, building quality, household resources, access to services etc) and the overall goal (reflecting the decisions to be made) was an assessment of the way in which those indicators contribute to (and reflect) SEP.

Methods

The Analytic Hierarchy Process

The AHP is based around a pairwise comparison procedure in which each condition is compared with every other condition and assigned a score in relation to the objective of interest (Saaty 1980).

Our approach to the pairwise comparison procedure was to first ask (oneself) which of the two conditions under comparison was likely to be better in indicating higher SEP (i.e. expected to have a higher score). The second stage (in which the score was assigned) was an assessment of the extent to which a household that possessed this condition was better-off (in terms of SEP) compared to one that didn't. For example, consider the scenario in which we compare household ownership of a mobile

phone and ownership of a television. One might expect that a television is a better indicator of household SEP than a mobile phone: the former is relatively high value, that is likely to be owned by few households, and typically only by those households who also have access to a source of power to run it. Now, imagine two households, both of which own a mobile phone, but only one of which also owns a television. Without any additional information, we ask ourselves how much 'better off' is the household with the television. Using the preference statements listed Table A5.2, we might expect (and others may have a different view) that a household with a television would be much to extremely better off than a household without (and therefore receive a value (from Table A5.2) of 8), given that the only additional information we have is that both households also own a mobile phone. This score can therefore be considered to represent the information that each factor holds in the assessment of SEP: ownership of a television contains very much more information with regard to a household's higher SEP than a mobile phone, whilst ownership of a mobile phone contains very much less information than the ownership of a television (1/8, or 0.125). This seems reasonable; mobile phones are now ubiquitous in the study area, can be purchased relatively cheaply and can be charged by small 'phone charging' businesses for minimal cost.

Consider a second pairwise comparison: the ownership of a sewing machine and a television. Both items represent a substantial cash investment for a household, but a sowing machine also represents a potential source of income, and may therefore be a better indicator of higher SEP. The pairwise comparison of television and sowing machine would therefore involve comparing two households, both of which own a television, only one of which owns a sewing machine. Without any other information, we ask ourselves how much better off is the household with a sewing machine than the household without, and decide that it is slightly better off (receiving a value of 3). Thus, compared to a television, a sewing machine provides slightly more information (3) about a household's higher SEP, whilst a television in comparison with a sewing machine provides slightly less (1/3, or 0.333) (but very much more than a mobile phone, and so on and so forth).

Comparisons proceeded in this way for each of the conditions (variables) in each domain (material wealth, access to services and household resources) (allowing comparison with the output from the domain based MCA conducted (automatically) as part of the MFA procedure) as well as for the global group of variables (allowing comparison with the overall MFA).

Table A5.2. Preference statements and values used in pairwise-comparisons (modified from Saaty (1980) to fit the inferential process): "how much better off is household *y* compared to household *x* given that it has condition *z*?"

Value	Description
1	No difference
2	Equal to slightly better off
3	Slightly better off
4	Slightly to moderately better off
5	Moderately better off
6	Moderately to much better off
7	Much better off
8	Much to extremely better off
9	Extremely better off

Weightings for each condition (variable) were derived from the output of this pairwise comparison procedure by normalising the eigenvector associated with the maximum eigenvalue of the resulting pairwise comparison matrix. The overall consistency of comparisons were assessed by calculating a 'consistency index' (CI) as the eigenvector of the normalised pairwise comparison matrix. The consistency index was then divided by a value generated by a random matrix (available from Saaty 1980) to derive a consistency ratio (CR). A CR less than 0.1 is generally considered as providing evidence that comparisons were highly consistent, but may be considered tolerable up to 0.2 (Saaty 1980).

The AHP procedure was implemented in Microsoft Excel (2010). All pairwise comparisons were conducted by the main author.

Following the derivation of weights for each condition, households were assigned a SEP score by multiplying the weights by a dummy indicator (0/1) for the households possession (or lack) of each condition.

The output from the AHP was compared with the first principal component from the MFA (described in Chapter 5). Total livestock value was excluded to allow direct comparison with the AHP procedure (which cannot (easily) deal with continuous measures). In addition, and to compare the MFA with a more typically used approach to deriving an index of SEP, we included all of the component variables into a single MCA (i.e. without their grouping into specific domains, as was performed for the MFA).

The MCA was performed using the *FactoMineR* package (Le et al. 2008) in R and compared to the index derived from MFA using the same set of variables.

Results

The resulting weights for the AHP are shown in Table A5.3. The full AHP had a CR of 0.11, suggesting a reasonably consistent degree of agreement.

The comparison of the output from the AHP with the output from the MFA (used as the index of SEP throughout this thesis) is shown in Figure A5.2. There was a very strong correlation between the two (Spearman's rho = 0.94, p = <0.001).

There was also a very strong positive correlation comparing the outputs from the MFA and MCA on all conditions (variables) (Spearman's rho = 0.96, p<0.001, Figure A5.3), suggesting that either approach could have been used.

Table A5.3. Condition (variable) weights derived from the AHP procedure

Domain/condition	Weight	Domain/condition	Weight
Household wealth		**Household services**	
Bed	0.009	Bed net	0.019
Mobile phone	0.011	Spring water	0.047
Cupboard	0.013	Well water	0.047
Sofa with cushions	0.022	Water treatment	0.073
Clock	0.023	Latrine	0.074
Watch	0.023	Borehole water	0.122
Torch	0.024	Piped	0.293
Radio	0.028	Access to power supply	0.325
Phone charger	0.040	**Household resources**	
Bicycle	0.051	Established > 5 years	0.045
Improved walls	0.058	More adults than children	0.073
Improved roof	0.074	> 4 adults	0.104
Improved floors	0.093	Male head	0.216
Television	0.135	Educated adult	0.231
Sewing machine	0.155	External income	0.332
Motorbike	0.242		

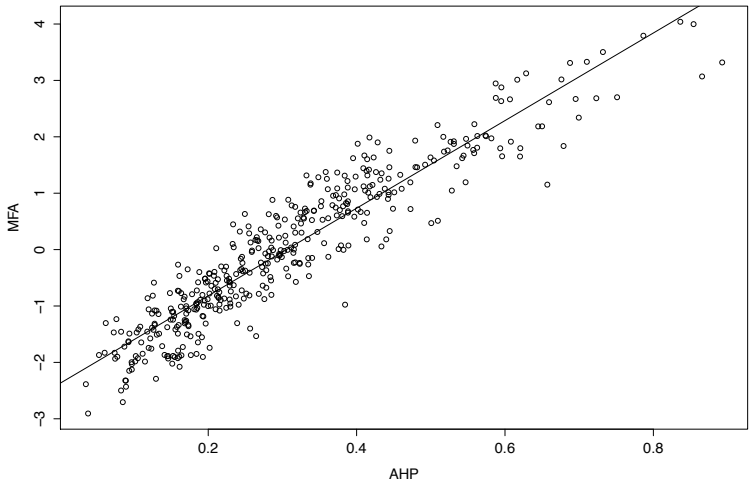


Figure A5.2. Relationship between **SEP index** derived using AHP on all conditions and the output from the MFA

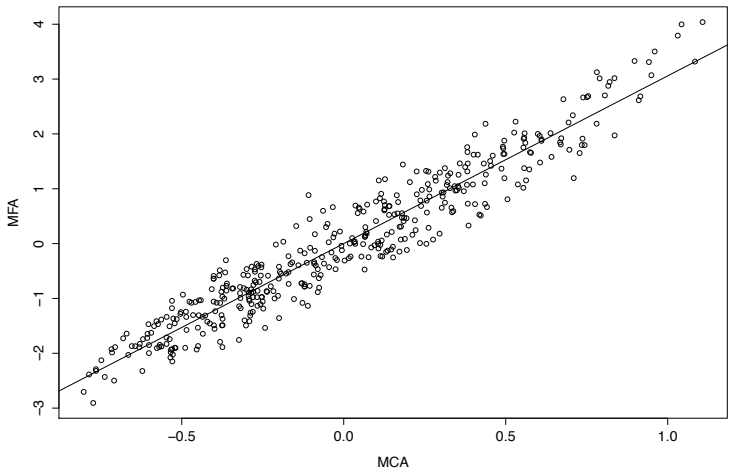


Figure A5.3. Comparison of domain-based MFA with full the MCA

Discussion and conclusions

There was a satisfyingly high level of agreement between the outputs from the AHP and the MFA, suggesting that the index of SEP (defined on the basis of MFA) used throughout this thesis is performing a reasonable job of approximating it, or at least our own subjective assessment of the influence of household conditions on SEP.

One major criticism of the validation approach used here is that the pairwise comparisons for the AHP were performed by a single person who also performed the MFA. It is possible that the outputs from the latter could have (inadvertently) biased the pairwise comparisons performed in the former. It

would of course be preferable for an uninvolved person or group to perform these comparisons, and ideally by one or more members of the community to which they are being applied. The collection of such data can be considered for the future, with a more formal comparison of the subjective with the objective measures.

To our knowledge, this is the first application of the AHP to the assignment of SEP. The AHP is a rigorous, straightforward, intuitive approach that, in this application, resulted in an output that is highly comparable to that derived using an objective approach. Further work is needed (and particularly to get an uninvolved party with local knowledge to perform the comparisons) but it could be a useful addition to the epidemiologist's tool box for assessing SEP in health surveys.

A5.4 Model diagnostics

Here, we present the variance inflation factors and exploration of residual spatial autocorrelation (RSA) of models described in Section 5.2.3 of the main text of Chapter 5.

Entamoeba

The final model describing relationship between *Entamoeba histolytica/dispar* and SEP with important covariates and hypothesised mediators:

```
summary(fit<-glmer(Entaemoba_histolytica ~ SEP + sex + rcs(age,4) +
HIV_Result + individual_latrine + (1 | HSID) + (1 | DIVID), data =
data, family = binomial))
```

Variance inflation factors:

SEP = 1.05; sex = 1.02; individual HIV result = 1.05; Individual latrine use = 1.12 (we do not estimate VIFs for those terms with restricted cubic splines (i.e. age in the above example)).

The Moran's I statistic (which has a positive valUe when positive spatial autocorrelation exists and a value of 0 when it does not) did not indicate RSA (Moran's I = 0.0039, p = 0.36).

Malaria

The final model describing relationship between *Plasmodium falciparum* and SEP with important covariates and hypothesised mediators:

```
summary(fit<-glmer(malaria ~ SEP +  age + urban_distance + pop_dens
+ HIV_Result + recent_antimalarial + (1 | HSID) + (1 | DIVID), data
= data, family = binomial))
```

Variance inflation factors:

SEP = 1.03; age = 1.04; urban distance = 1.01; population density = 1.03; individual HIV result = 1.03; recent antimalarial=1.01

The Moran's I statistic did not indicate RSA (Moran's I = 0.005, p = 0.29).

TB

The final model describing relationship between TB and SEP with important covariates and hypothesised mediators:

```
summary(fit<-glmer(Quantiferon_Result ~ SEP + age + HIV_Result + (1
| HSID), data = data, family = binomial))
```

Variance inflation factors:

SEP = 1.00; age = 1.10; HIV result = 1.01

The Moran's I statistic did not indicate RSA (Moran's I = 0.009, p = 0.11).

Hookworm

The final model describing relationship between hookworm and SEP with important covariates and hypothesised mediators:

```
summary(fit<-glmer(hookworm ~ SEP + sex + age + I(age^2) + ndvi +
I(ndvi^2) + luo +  HIV_Result + individual_latrine + (1 | HSID) + (1
| DIVID), data = data, family = binomial))
```

Variance inflation factors:

SEP= 1.06; sex = 1.03; age = 2.18; age^2 = 2.07; NDVI = 1.12; NDVI^2 = 1.11; Luo = 1.02; HIV result = 1.03; individual latrine = 1.12.

The Moran's I statistic did not indicate RSA (Moran's I = 0.009, p = 0.09).

Trichuris

The full model for *Trichuris* (including SEP and all hypothesised predictors) showed some evidence of RSA (Moran's I = 0.015, p =0.01). Given that there was substantial overlap between the distributions of household level HIV and *Trichuris* infection (observed using the risk surfaces developed using the KDE procedure in Chapter 4), we explored whether the inclusion of household level predictors of HIV would remove the residual spatial autocorrelation. These covariates were i. presence of at least person with HIV in the household; ii. prevalence of HIV in the household; iii. the count of HIV positive individuals in the household.

All of covariates were significant when included (individually) in the full model (i. OR = 1.77 (95% CI 1.06 – 2.96) ii. OR = 5.84 (95% CI 1.26 -27.1) iii. OR = 1.50 (95% CI 1.11 – 2.05)), and all except the prevalence of HIV in the household resulted in the p-value for the Moran's I statistic being (marginally) greater than the 5% cut-off (i. Moran's I = 0.011, p = 0.06 ii. Moran's I = 0.012, p = 0.037; iii. Moran's I =0.011, p = 0.051). Since the variable describing the count of HIV positive individuals in a household resulted in a better fit than the presence of at least one HIV positive person (AIC 1153 vs. 1155), we took this variable forward.

Using the model selection procedure described in the main text, we derived the following final model describing the relationship between *Trichuris trichiura* and SEP and important covariates, hypothesised confounders and mediators:

```
summary(fit<-glmer(trichuris ~ SEP + sex + age + pop_dens + lst_max
+ luo + individual_latrine + HIV_Result + HIV_count + (1 | HSID) +
(1 | DIVID), data = data, family = binomial))
```

Variance inflation factors:

SEP = 1.13; sex = 1.03; age = 1.04; population density = 1.14; maximum LST = 1.03; Luo = 1.03; HIV result = 1.08; individual latrine = 1.10; HIV count = 1.21.

The exclusion of some of the household level predictors from this model (as part of the model selection procedure) resulted in the p-value of the Moran's I test for this final model being just below the 5% cut-off (Moran's I = 0.011, p-value = 0.049). However, there was little difference in the final OR for the effect of SEP on *Trichuris* infection, or its confidence interval, compared to the full model (1.53 (95% CI 1.13 – 2.08) vs 1.50 (95% CI 1.11 – 2.05), so this was considered to be of little consequence for the main aims of this study.

Ascaris

Adopting the model building strategy described in the main text, the final model describing the relationship between *Ascaris* and SEP with important covariates and hypothesised mediators, was:

```
summary(fit<-glmer(ascaris ~ SEP + rcs(age,3) + luo + HIV_Result +
individual_latrine + HIV_count + (1 | HSID), data = data, family =
binomial))
```

However, the Moran's I statistic indicated strong evidence of residual spatial autocorrelation (RSA) (Moran's I = 0.022, p =<0.001) (even with HIV count in the model, which was included given its observed importance for *Trichuris*, and the observed overlap (in Chapter 4) between the spatial distribution of household risk of *Trichuris* and *Ascaris*), therefore we refitted the model with the x and y co-ordinates for each household in order to remove the effect of broad spatial trends:

```
summary(fit<-glmer(ascaris ~ SEP + rcs(age,3) + HIV_Result +
individual_latrine + HIV_count + x_coord + y_coord + (1 | HSID),
data = data, family = binomial))
```

There was still some indication of RSA in the resulting Moran's I, although the p-value is above the 5% cut-off (Moran's I = 0.01, p =0.053). The significant effect of household membership of the Luo tribe (OR = 3.4 (95% CI 1.5 – 7.5)) was no longer significant and was removed following the inclusion of the x and y co-ordinates.

Variance inflation factors for the final model:

SEP = 1.14; HIV result (individual) = 1.07; HIV count (household) = 1.26; individual latrine use = 1.11; latitude = 1.53; longitude = 1.44.

Use of antimalarials

The final model describing the relationship between the use of antimalarials and SEP with important covariates was:

```
summary(fit<-glmer(recent_antimalarial ~ rcs(SEP,3) + age + I(age^2)
+ sex + (1 | HSID) + (1 | DIVID), data = data, family = binomial))
```

Variance inflation factors:

Age = 2.20; age^2 = 2.19; sex = 1.02

There was no evidence of residual spatial autocorrelation (Moran's I = 0.00057, p = 0.66).

Use of antibiotics

The final model describing the relationship between the use of antibiotics and SEP with important covariates was:

```
summary(fit<-glmer(recent_antibiotics ~ SEP +  sex + (1 | HSID) + (1
| DIVID), data = data, family = binomial))
```

Variance inflation factors:

SEP = 1.00; sex = 1.00

There was no evidence of residual spatial autocorrelation (Moran's I = -0.001, p = 0.93).

Use of anti-inflammatories

The final model describing the relationship between the use of anti-inflammatories and SEP with important covariates was:

```
summary(fit<-glmer(recent_antiinflammatory ~ SEP + sex +
urban_distance +(1 | HSID) + (1 | DIVID), data = data, family =
binomial))
```

Variance inflation factors:

SEP = 1.00; sex = 1.00; urban distance = 1.00

There was no evidence of residual spatial autocorrelation (Moran's I = 0.007, p = 0.16)

Appendix to Chapter 6

A6.1 Univariable analysis

Whilst the model selected by purposeful selection in chapter 6 was derived from the starting point of a full model (i.e. no pre-selection was used), we include the univariable results here to assist with interpretation and to allow cross-reference with the effects in the multivariable model.

Tables A6.1, A6.3 and A6.5 show the outputs from the univariable analysis for predictors of detection for hookworm, *Ascaris* and *Trichuris*, respectively. Tables A6.2, A6.4 and A6.6 show the outputs from the univariable analysis for occupancy. We show two results for each predictor in each of these tables: one in which the there are no predictors on detection; one in which all predictors for detection (selected by purposeful selection) are included.

Table A6.1. Univariable analysis of predictors for hookworm detection.

	Detection				
	Intercept (SE)	**Parameter (SE)**	**p-value**	**AIC**	**ΔAIC**
Null model	-0.26 (0.06)	-	-	2463	-
Individual level					
Age	-0.25 (0.06)	0.27 (0.054)	<0.001	2440	-35
HIV infection	-0.21 (0.06)	-1.14 (0.27)	<0.001	2445	-18
Male	-0.38 (0.07)	0.27 (0.10)	0.01	2459	-4
Anaemia	-0.20 (0.06)	-0.23 (0.12)	0.05	2462	-1
Household level					
Socioeconomic status	-0.19 (0.06)	-0.35 (0.06)	<0.001	2426	-37
Latitude	-0.31 (0.06)	0.33 (0.06)	<0.001	2434	-29
HIV infection in household	-0.25 (0.06)	-0.17 (0.06)	0.001	2454	-9
Longitude	-0.27 (0.06)	0.16 (0.05)	0.002	2456	-7
Local population density	-0.26 (0.06)	-0.18 (0.06)	0.005	2458	-5
Land surface temperature	-0.25 (0.06)	0.15 (0.06)	0.01	2459	-4
NDVI	-0.27 (0.06)	0.10 (0.06)	0.16	2463	0
Household count	-0.22 (0.06)	-0.054 (0.05)	0.24	2464	+1

Models contain no predictors of occupancy, and a single predictor of detection ($x_{i,j}$) i.e:
$logit(\psi_i) = \beta_{occupancy}$
$logit(p_{i,j}) = \beta_{detection} + \beta_1 * x_{i,j}$

Table A6.2.: Univariable (left) and multivariable (right) analysis of predictors for hookworm occupancy. The multivariable analysis includes all predictors of detection selected by purposeful selection and one of each of the predictors of occupancy

	Intercept[1] (SE)	Parameter[1] (SE)	p - value	AIC	Δ AIC	Intercept[2] (SE)	Parameter[2] (SE)	p - value	AIC	Δ AIC
Null model	1.42 (0.16)	-	-	2463	-	1.54 (0.18)	-	-	2332	
NDVI	1.47 (0.17)	0.48 (0.15)	0.001	2455	-8	1.61 (0.20)	0.54 (0.17)	0.001	2323	-9
Latitude	1.50 (0.18)	0.53 (0.20)	0.006	2456	-7	1.51 (0.18)	0.34 (0.19)	0.080	2331	-1
Pop. density	1.45 (0.17)	-0.36 (0.14)	0.01	2459	-4	1.56 (0.19)	-0.36 (0.15)	0.018	2329	-3
Longitude	1.43 (0.16)	0.25 (0.16)	0.11	2463	0	1.52 (0.18)	0.14 (0.17)	0.40	2333	+1
HIV	1.54 (0.35)	-0.48 (0.35)	0.17	2463	0	1.52 (0.18)	0.14 (0.17)	0.40	2333	+1
SES	1.46 (0.17)	-0.18 (0.15)	0.22	2464	+1	1.54 (0.18)	0.043 (0.18)	0.81	2334	+2
LST	1.43 (0.16)	0.13 (0.15)	0.37	2465	+2	1.54 (0.18)	0.11 (0.18)	0.54	2334	+2
Count	1.39 (0.16)	0.11 (0.16)	0.50	2465	+2	1.50 (0.18)	0.16 (0.18)	0.38	2333	+1

1 $logit(\psi_i) = \beta_{occupancy} + \beta_1 * x_i$
$logit(p_{i,j}) = \beta_{detection}$

2 $logit(\psi_i) = \beta_{occupancy} + \beta_1 * x_i$
$logit(p_{i,j}) = \beta_{detection} + \beta_1 * Age_{i,j} + \beta_2 * Age^2{}_{i,j} + \beta_3 * Gender_{i,j} + \beta_5 * HIV\ (individual)_{i,j} + \beta_6 * SES_{i,j} + \beta_7 * Northing_{i,j} + \beta_8 * Household\ count_{i,j}$

Table A6.3. Univariable analysis of predictors for *Ascaris* detection.

	Detection				
	Intercept	**Parameter (SE)**	**p-value**	**AIC**	**ΔAIC**
Null model	-1.19 (0.12)	-	-	1077	-
Individual level					
Age	-1.28 (0.12)	-0.54 (0.11)	<0.001	1052	-25
HIV infection	-1.14 (0.12)	-0.72 (0.42)	0.09	1076	-1
Male	-1.12 (0.14)	-0.16 (0.18)	0.38	1079	+2
Anaemia	-1.20 (0.13)	0.033 (0.20)	0.87	1079	+2
Household level					
Socioeconomic status	-1.11 (0.12)	-0.32 (0.10)	0.0017	1069	-8
Longitude	-1.27 (0.12)	0.26 (0.09)	0.0059	1071	-6
NDVI	-1.21 (0.12)	0.20 (0.11)	0.066	1076	-1
Local population density	-1.16 (0.12)	-0.11 (0.09)	0.22	1077	0
Household count	-1.12 (0.15)	-0.073 (0.11)	0.50	1079	+2
Land surface temperature	-1.18 (0.12)	0.020 (0.08)	0.80	1079	+2
HIV infection in household	-1.19 (0.12)	0.01 (0.08)	0.9	1079	+2
Latitude	-1.19 (0.13)	-0.007 (0.12)	0.95	1079	+2

$logit(\psi_i) = \beta_{occupancy}$
$logit(p_{i,j}) = \beta_{detection} + \beta_1 * x_{i,j}$

Table A6.4.: Univariable (left) and multivariable (right) analysis of predictors for *Ascaris* occupancy. Multivariable analysis includes all predictors of detection selected by purposeful selection and one of each predictor of occupancy

	Intercept[1] (SE)	**Parameter[1] (SE)**	**p - value**	**AIC**	**Δ AIC**	**Intercept[2] (SE)**	**Parameter[2] (SE)**	**p - value**	**AIC**	**Δ AIC**
Null model	-0.56 (0.16)	-	-	1077	-	-0.47 (0.16)	-	-	1041	-
Latitude	-0.58 (0.16)	-0.57 (0.16)	<0.001	1064	-13	-0.49 (0.17)	-0.62 (0.16)	<0.001	1026	-15
Longitude	-0.58 (0.16)	0.25 (0.14)	0.07	1076	-1	-0.50 (0.16)	0.21 (0.14)	0.13	1040	-1
HIV	-0.69 (0.17)	0.53 (0.33)	0.10	1077	0	-0.61 (0.18)	0.49 (0.34)	0.15	1041	0
LST	-0.57 (0.16)	-0.18 (0.13)	0.17	1077	0	-0.48 (0.16)	-0.19 (0.14)	0.16	1041	0
NDVI	-0.57 (0.16)	0.16 (0.14)	0.27	1078	+1	0.49 (0.16)	0.08 (0.17)	0.64	1042	0
Count	-0.62 (0.16)	0.13 (0.13)	0.31	1078	+1	-0.52 (0.17)	0.11 (0.14)	0.41	1042	+1
Pop. dens	-0.56 (0.16)	0.12 (0.16)	0.46	1079	+2	-0.47 (0.16)	0.20 (0.17)	0.24	1041	0
SES	-0.56 (0.16)	-0.03 (0.14)	0.81	1079	+2	-0.45 (0.17)	0.26 (0.17)	0.12	1040	-1

1 $logit(\psi_i) = \beta_{occupancy} + \beta_1 * x_i$
$logit(p_{i,j}) = \beta_{detection}$

2 $logit(\psi_i) = \beta_{occupancy} + \beta_1 * x_i$
$logit(p_{i,j}) = \beta_{detection} + \beta_1 * Age_{i,j} + \beta_2 * HIV\ (household)_{i,j} + \beta_3 * HIV_{i,j} + \beta_4 * SES_{i,j} + \beta_5 * NDVI_{i,j}$

Table A6.5. Univariable analysis of predictors for *Trichuris* detection.

	Detection				
	Intercept	**Parameter (SE)**	**p-value**	**AIC**	**ΔAIC**
Null model	-1.18 (0.11)	-	-	1278	
Individual level					
Male	-1.01 (0.12)	-0.42 (0.16)	0.01	1273	-5
Age	-1.18 (0.10)	-0.20 (0.08)	0.02	1274	-4
HIV infection	-1.10 (0.11)	-0.76 (0.36)	0.04	1275	-3
Anaemia	-1.26 (0.12)	0.25 (0.18)	0.16	1278	0
Household level					
Latitude	-2.17 (0.13)	-1.10 (0.11)	<0.001	1190	-88
Land surface temperature	-1.34 (0.11)	-0.41 (0.08)	<0.001	1250	-28
Local population density	-1.41 (0.12)	0.41 (0.08)	<0.001	1254	-24
Longitude	-1.35 (0.12)	-0.32 (0.09)	<0.001	1267	-11
HIV infection in household	-1.36 (0.13)	0.17 (0.06)	0.005	1272	-6
Socioeconomic status	-1.27 (0.12)	0.18 (0.10)	0.07	1277	-1
Household count	-1.15 (0.17)	0.061 (0.07)	0.38	1279	+1
NDVI	-1.18 (0.11)	-0.059 (0.08)	0.46	1280	+2

$logit(\psi_i) = \beta_{occupancy}$
$logit(p_{i,j}) = \beta_{detection} + \beta_1 * x_{i,j}$

Table A6.6: Univariable (left) and multivariable (right) analysis of predictors for *Trichuris* occupancy. Multivariable analysis includes all predictors of detection selected by purposeful selection and one of each predictor of occupancy

	Intercept[1] (SE)	Parameter[1] (SE)	p - value	AIC	Δ AIC	Intercept[2] (SE)	Parameter[2] (SE)	p - value	AIC	Δ AIC
Null model	-0.17 (0.16)	-	-	1278	-	0.20 (0.22)	-	-	1219	-
Latitude	-0.12 (0.26)	-2.41 (0.45)	<0.001	1166	-112	0.23 (0.32)	-2.86 (0.56)	<0.001	1121	
HIV	-0.5 (0.17)	1.5 (0.40)	<0.001	1261	-17	-0.26 (0.42)	1.26 (0.42)	0.003	1207	
Longitude	-0.16 (0.16)	-0.46 (0.14)	0.001	1268	-10	0.12 (0.21)	-0.31 (0.15)	0.04	1206	
Pop. dens	-0.12 (0.17)	0.76 (0.27)	0.004	1268	-10	0.10 (0.21)	0.50 (0.32)	0.12	1207	
Count	-0.29 (0.16)	0.29 (0.13)	0.027	1275	-3	0.06 (0.21)	0.28 (0.11)	0.11	1208	
LST	-0.17 (0.16)	-0.32 (0.19)	0.08	1276	-2	0.25 (0.15)	0.076 (0.15)	0.60	1210	
SES	-0.19 (0.16)	0.15 (0.14)	0.30	1279	+1	0.17 (0.22)	0.21 (0.18)	0.24	1209	
NDVI	-0.17 (0.16)	-0.08 (0.13)	0.54	1280	+2	0.20 (0.22)	-0.086 (0.15)	0.56	1210	

1 $logit(\psi_i) = \beta_{occupancy} + \beta_1 * x_i$
$logit(p_{i,j}) = \beta_{detection}$

2 $logit(\psi_i) = \beta_{occupancy} + \beta_1 * x_i$
$logit(p_{i,j}) = \beta_{detection} + \beta_1 * Age_{i,j} + \beta_2 * HIV\ (household)_{i,j} + \beta_3 * HIV_{i,j} + \beta_4 * Gender_{i,j} + \beta_5 * Count_{i,j} + \beta_6 * \quad Pop.density_{i,j} + \beta_7 * Pop.density^2{}_{i,j} + \beta_8 * LST_{i,j}$

A6.2 Meta-analysis of the effect of HIV on helminth prevalence

We reviewed relevant literature in Web of Science using the search terms [(HIV AND (Ascaris OR Trichuris OR hookworm))]. Articles were reviewed and studies which reported prevalence of infection for any of the STH species in HIV positive individuals and HIV negative controls were selected for inclusion. Given the surprisingly small number of studies identified, we specified no selection criteria in terms of quality of studies, and substantial differences in the choice of controls were reported in all. None of the studies reported the anthelmintic treatment history in either the HIV infected or HIV negative group. Some of the studies included individuals receiving anti-retrovirals, and in some HIV positive individuals were in advanced stages of AIDS.

A description of the studies, together with reported prevalence is given in Table A6.7

To compare the prevalence for each helminth infection in the two groups (HIV +ve and HIV –ve) we followed that used by Poulin (1996) and Brooker et al. (2004a) who both compared prevalence of helminth infection according to sex. Differences in prevalence were calculated using the following formula:

$$(p_{hiv+} - p_{hiv-})(J), where\ J = 1 - \frac{3}{4(N_{hiv+} + N_{hiv1} - 2) - 1}$$

The difference between the prevalence in HIV positive (p_{hiv+}) and HIV negative (p_{hiv-}) is weighted by *J*, a correction by small sample sizes (N): the larger the sample size, the closer J is to one and therefore the greater the influence the study has (Poulin 1996).

If there are no differences in STH prevalence between those people with and without HIV, differences in prevalence are expected to be normally distributed around 0. A one-group, two-tailed t-test was used to compare the resulting differences with an expected mean of 0.

Findings

HIV infected individuals had a lower average prevalence than people without for all three of the STH, although only in the case of hookworm was this difference significant (mean difference hookworm = -4.6% (95% CI -8.9%, -0.4%); mean difference *Ascaris* = -3.8% (95% CI -7.7%, 0.2%); mean difference *Trichuris* = -1.6% (95% CI -6.8%, 3.7%) (Table A6.7 and Figure A6.1).

Table A6.7 Summary of identified literature describing difference in STH prevalence between HIV + and HIV- individuals (continued on next page).

Study description	Parasite	HIV+		HIV-	
		n/N	% (95% CI)	n/N	% (95% CI)
Patients attending clinic (HIV+); Randomly selected household member of HIV+ (HIV-). All ages and gender. No matching. Equatorial Guinea. (Roka et al. 2013)	Trichuris	124/503	23.4 (19.8 – 23.0)	30/131	22.9 (15.7 – 30.1)
	Ascaris	63/503	11.9 (9.1 – 14.6)	19/131	14.5 (8.5 – 20.5)
	Hookworm	11//503	2.1 (0.9 – 3.3)	4/131	3.1 (0.1 – 6.0)
Patients attending clinic (HIV+); Randomly selected household member of HIV+ (HIV). All ages and gender. No matching. Equatorial Guinea (an island of). (Roka et al. 2012)	Trichuris	134/260	51.5 (45.3 – 57.8)	21/50	42.0 (14.5 – 41.4)
	Ascaris	50/260	19.2 (14.2 – 24.2)	14/50	28.0 (14.5 – 41.4)
	Hookworm	11/260	4.2 (1.6 – 6.9)	5/50	10.0 (3.3 – 21.8)
Patients attending clinic (HIV+ and HIV-); Selection procedure not described in detail. Ghana. (Boaitey et al. 2012)	Hookworm	5/500	1.0 (0.3 – 2.5)	2/300	0.6 (0.1 – 2.6)
Patients attending 3 outpatients clinic (HIV+ and HIV-): Selection procedure not described in detail: patients aged 4 to 64, and of male and female gender. No matching. Nigeria. (Sanyaolu et al. 2011)	Trichuris	3/65	4.6 (1.2 – 13.8)	40/1015	3.9 (2.9 – 5.4)
	Ascaris	5/65	7.7 (2.9 – 17.8)	97/1015	9.6 (7.9 – 11.6)
	Hookworm	3/65	4.6 (1.2 – 13.8)	18/1015	1.8 (1.1 – 2.8)
Adults attending HIV support sessions (HIV+, ARV naïve); healthy adults accompanying them (HIV-); recruited as part of study of deworming; South Africa. (Mkhize-Kwitshana et al. 2011)	Ascaris	51/124	41.1 (32.5 – 50.3)	18/39	46.1 (30.4 – 62.6)
Individuals with pulmonary TB (sputum smear or culture positive) aged more than 5 years. Tanzania. (Range et al. 2007)	Hookworm	28/232	12.1 (8.3 – 17.1)	75/300	25.0 (20.3 – 30.4)
Individuals suspected of pulmonary TB but negative on sputum smear or culture, aged more than 5 years. Tanzania. (Range et al. 2007)	Hookworm	5/77	6.5 (2.4 – 15.2)	11/46	23.9 (13.1 – 39.1)
Patients attending HIV counselling and testing service (HIV+ (ARV naïve), HIV-). Patients of all ages and gender. No matching. Malawi. (Hosseinipour et al. 2007)	Ascaris	8/266	3.0 (1.4 – 6.1)	3/123	2.4 (0.6 – 7.5)
	Hookworm	27/266	10.2 (6.9 – 14.6)	35/123	28.5 (20.9 – 37.4)
Cross-sectional study of adults from 5 villages (HIV+ (ARV naïve); HIV-). Tanzania. (Nielsen et al. 2006)	Trichuris	4/46	8.7 (2.8 – 21.7)	41/596	6.9 (0.05 – 0.093)
	Ascaris	0/46	0 (0 – 9.6)	8/596	1.3 (0.6 – 2.7)
	Hookworm	29/46	63.0 (47.5 – 76.4)	456/596	76.5 (72.9 – 79.8)
Hospitalised HIV/AIDs patients (HIV+, 73% ARV naive) ; clinically healthy volunteers without high risk practices (HIV-ve). Aged 18 – 76. Brazil. (Silva et al. 2005)	Trichuris	1/100	1.0 (0.05 – 6.2)	1/85	1.2 (0.6 – 7.3)
	Ascaris	0/100	0 (0 – 4.6)	1/85	1.2 (0.6 – 7.3)
	Hookworm	4/100	4 (1.3 – 10.5)	6/85	7.1 (2.9 – 15.3)

Patients with AIDs (HIV+, unclear if on ARVs) and healthy controls (HIV–ve, unclear if serologically confirmed as HIV free). Ethiopia. (Hailemariam et al. 2004)	Trichuris	5/78	6.4 (2.4 – 15.0)	2/26	7.7 (1.3 – 26.6)
	Ascaris	24/78	30.8 (21.1 – 42.4)	6/26	23.1 (9.8 – 44.1)
	Hookworm	2/78	2.6 (0.4 – 9.8)	1/26	3.8 (0.2 – 21.6)
Cross-sectional survey of adults living (and working) on sugar plantation (HIV+ (ARV naïve); HIV). Ethiopia. (Fontanet et al. 2000)	Trichuris	6/52	11.5 (4.8 – 24.1)	235/1187	19.7 (17.6 – 22.2)
	Ascaris	6/52	11.5 (4.8 – 24.1)	267/1187	22.5 (20.2 – 25.0)
	Hookworm	9/52	17.3 (8.7 – 30.8)	286/1187	24.1 (21.7 – 26.7)
Adults attending hospital (presenting history unclear) (HIV+ (no information on ARV status) and HIV-). Brazil. (Marchi Blatt & Cantos 2003)	Trichuris	1/211	0.5 (0 – 3.0)	3/213	1.4 (0.3 – 4.4)
	Ascaris	5/211	2.4 (0.8 – 5.7)	11/213	5.2 (2.7 – 9.3)
	Hookworm	0/211	0 (0 – 2.2)	3/213	0.9 (0.1 – 3.7)
Individuals attending AIDs clinic (HIV+ (no information on ARV status)); controls selected randomly from faecal examinations performed on HIV negative individuals as part of hospital routine diagnostics (HIV-), and therefore likely to be biased. (Feitosa et al. 2001)	Trichuris	19/365	5.2 (3.2 – 8.1)	291/5243	5.6 (5.0 – 6.2)
	Ascaris	43/365	11.8 (8.7 – 15.6)	495/5243	9.4 (8.7 – 10.3)
	Hookworm	16/365	4.4 (2.6 – 7.2)	198/5243	3.8 (3.3 – 4.3)
Individuals with diarrhoea in hospital and prison clinic (HIV+ (ARC naïve) and HIV-). Aged 1 to 71 years. No matching. Honduras. (Lindo et al. 1998)	Trichuris	11/52	21.2 (11.6 – 35.1)	19/48	39.6 (26.1 – 54.7)
	Ascaris	1/52	1.9 (0.1 – 11.6)	10/48	20.8 (11.0 – 35.4)
	Hookworm	9/52	17.3 (8.7 – 30.8)	4/48	8.3 (2.7 – 20.9)
Individuals with diarrhoea (HIV+ (ARC naïve) and HIV-). No matching. Tanzania. (Morales et al. 1995)	Ascaris	4/112	3.6 (1.2 – 9.4)	25/239	10.5 (7.0 – 15.2)
	Hookworm	13/112	11.6 (6.6 – 19.4)	31/239	13.0 (9.1 – 18.1)

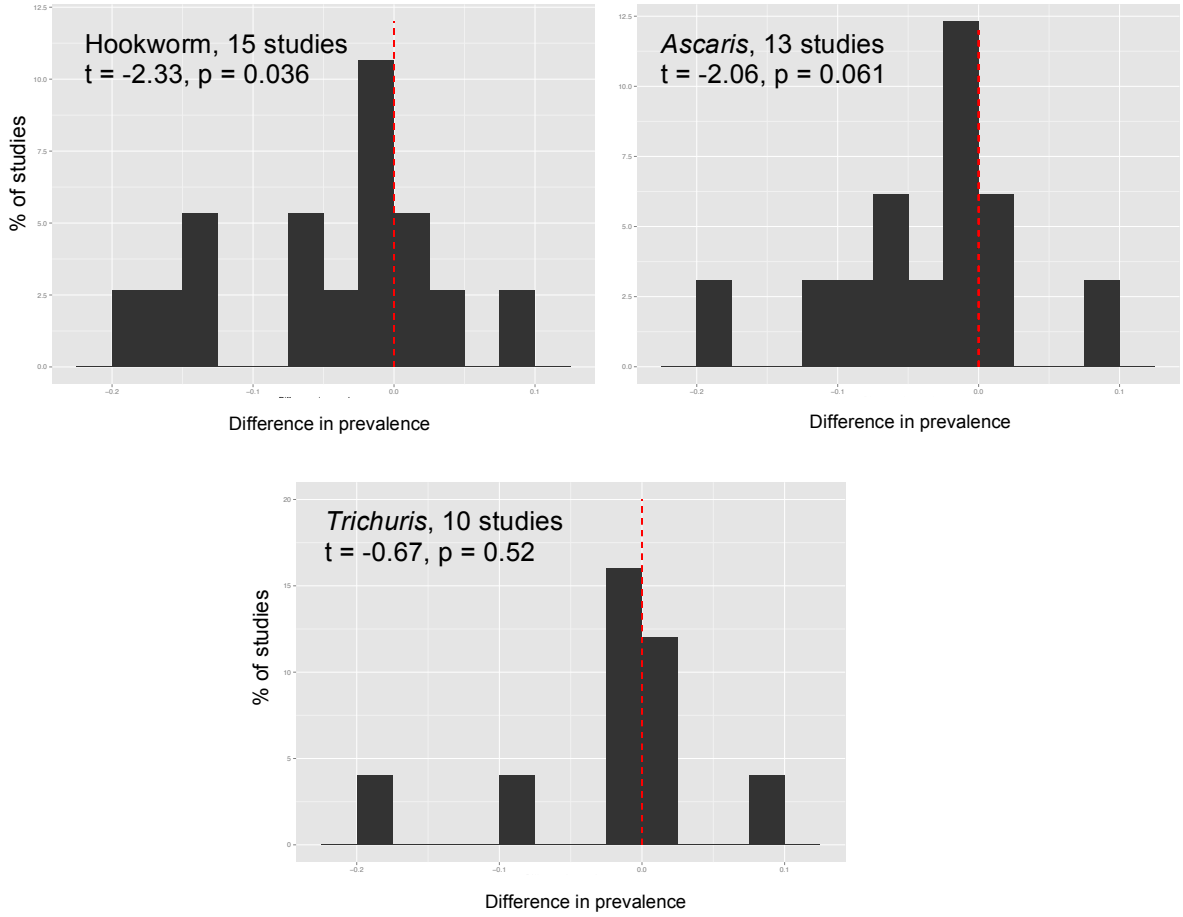


Figure A6.1. Summary of difference in prevalence between HIV+ and HIV- people (x axis) and the percentage of studies reporting difference (y axis). Red dashed line is at 0.

Appendix to Chapter 7

A7.1 Specifying the autocovariate

We estimated the autocovariate term for each household (ac_i) as a spatial term $s(y_i)$ where *y* values at location *j* in the neighbourhood of *i* are summarised as a distance weighted average. The data point *i* is excluded from the calculation (Augustin et al. 1996):

$$ac_i = s(y_i) = \frac{\sum_{j \neq i} w_j y_i}{\sum w_j} \text{ where } w_j = d_{ij}^{-1}$$

The following R code was used to derive the number of neighbours each household has, identifies those neighbours and estimates the distance between them (adapted from (Burton et al. 2012)).

```
library(spdep)
HS_p <- read.csv("...") # Household co-ordinates
nsite <- dim(HS_p)[1]
coords <- as.matrix(cbind(HS_p$POINT_X/1000,HS_p$POINT_Y/1000))
n.dist <- 15
nb <- dnearneigh(coords,0,n.dist)
nn <- rep(0,nsite)
for (n in 1:nsite) nn[n] <- length(nb[[n]])
NN <- matrix(rep(0,nsite*max(nn)),nrow=nsite)
for (n in 1:nsite)
{
  NN[n,] <- append(as.vector(nb[[n]]),rep(NA,max(nn)-nn[n]))
}
nn <- rep(0,nsite)
for (n in 1:nsite) nn[n] <-  ifelse( sum(nb[[n]])>0, length(nb[[n]]), 0)

nbd <- nbdists(nb, coords)
D <- matrix(0,nrow=nsite, ncol=max(nn))
for(n in 1:nsite)
{
  D[n,] <- append(as.vector(nbd[[n]]),rep(NA,max(nn)-nn[n]))
}
```

The following code derived the 5 nearest neighbours around each household.

```
XX = D
YY = NN
p <- matrix(NA,nrow= nsite, ncol=5)
l <- matrix(NA,nrow= nsite, ncol=5)

for(k in 1: nsite){
  p[k,1] = YY[k,which.min(XX[k,])]
  l[k,1] = XX[k,which.min(XX[k,])]
  XX[k,which.min(XX[k,])]=NA
  p[k,2] = YY[k,which.min(XX[k,])]
  l[k,2] = XX[k,which.min(XX[k,])]
  XX[k,which.min(XX[k,])]=NA
  p[k,3] = YY[k,which.min(XX[k,])]
  l[k,3] = XX[k,which.min(XX[k,])]
  XX[k,which.min(XX[k,])]=NA
  p[k,4] = YY[k,which.min(XX[k,])]
```

```
  l[k,4] = XX[k,which.min(XX[k,])]
  XX[k,which.min(XX[k,])]=NA
  p[k,5] = YY[k,which.min(XX[k,])]
  l[k,5] = XX[k,which.min(XX[k,])]
}

D = l
NN = p
```

A7.2 Hierarchical community model

The following (shortened) R code (adapted from Kery & Royle (2008) and Burton et al. (2012)) performs the multispecies hierarchical community model, estimates the Bayesian p-value, derives household prevalence estimates and estimates within household species richness.

```
sink("model_cov.txt")
cat("
    model {
    # priors
    p0 ~ dunif(0,1)
    psi0 ~ dunif(0,1)
    sigmap ~ dunif(0,5)
    sigmapsi ~ dunif(0,5)
    rho ~ dunif(-1,1)

    taup <- (1/(sigmap*sigmap))
    taupsi <- (1/(sigmapsi*sigmapsi))
    mup <- log(p0/(1-p0))
    mupsi <- log(psi0/(1-psi0))
    var.eta <- taup/(1.-pow(rho,2))

    mu.delta ~ dnorm(0, 0.01)
    sigma.delta ~ dunif(0,5)
    tau.delta <- (1/sigma.delta*sigma.delta)

    mu.betalpsi1 ~ dnorm(0,0.01)
    tau.betalpsi1 <- pow(sd.betapsi1,-2)
    sd.betapsi1 ~ dunif(0,3)
    ....
    mu.betalpsi8 ~ dnorm(0,0.01)
    tau.betalpsi8 <- pow(sd.betapsi8,-2)
    sd.betapsi8 ~ dunif(0,3)

    mu.betalp1 ~ dnorm(0,0.01)
    tau.betalp1 <- pow(sd.betap1,-2)
    sd.betap1 ~ dunif(0,3)
    ....
    mu.betalp9 ~ dnorm(0,0.01)
    tau.betalp9 <- pow(sd.betap9,-2)
    sd.betap9 ~ dunif(0,3)

    for(k in 1:(nspec)){
    # Randoms
    betalpsi0[k] ~ dnorm(mupsi, taupsi)
    betalpsi1[k] ~ dnorm(mu.betalpsi1, tau.betalpsi1)
    ....
    betalpsi8[k] ~ dnorm(mu.betalpsi8, tau.betalpsi8)
    delta[k] ~ dnorm(mu.delta, tau.delta)

    mu.lp[k] <- mup + (rho*sigmap/sigmapsi)*(betalpsi0[k]-mupsi)
```

```
    betalp0[k] ~ dnorm(mu.lp[k], var.eta)
    betalp1[k] ~ dnorm(mu.betalp1, tau.betalp1)
    ....
    betalp9[k] ~ dnorm(mu.betalp9, tau.betalp9)
    }

    # Likelihood
    # Ecological model for true occurrence (process model)
    for(k in 1:(nspec)){
    for (i in 1:nsite) {
    # derive neighbourhood infection status for autocovariate.
    x[i,k,1] <- 0
    for (g in 1:5){
    x[i,k,g+1] <- x[i,k,g] + z[NN[i,g],k]/D[i,g]
    }

    psi[i,k] <- 1/(1+exp(-lpsi.lim[i,k]))
    lpsi.lim[i,k] <- min(999, max(-999, lpsi[i,k]))
     lpsi[i,k] <- betalpsi0[k] + betalpsi1[k]*number[i] +
     betalpsi2[k]*lstmax[i] + betalpsi3[k]*ndvi[i] +
     betalpsi4[k]*popdens[i] + betalpsi5[k]*cows[i] + betalpsi6[k]*grass[i]
     + betalpsi7[k]*ses[i] + betalpsi8[k]*surface[i] +
     delta[k]*(x[i,k,5+1]/5)
    z[i,k] ~ dbern(psi[i,k])

    for(j in 1:nrep){
     logit(p[i,j,k]) <- betalp0[k] + betalp1[k]*age_i[i,j] +
     betalp2[k]*hiv_i[i,j] + betalp3[k]*sex_i[i,j] +
     betalp4[k]*milk_i[i,j] + betalp5[k]*ses_i[i,j] +
     betalp6[k]*number_i[i,j] + betalp7[k]*latrine_i[i,j] +
    betalp8[k]*vita_i[i,j] + betalp9[k]*iron_i[i,j]
    p.eff[i,j,k] <- z[i,k]*p[i,j,k]
    Y[i,j,k] ~ dbern(p.eff[i,j,k])
    Ynew[i,j,k] ~ dbern(p.eff[i,j,k])

    #Create simulated dataset to calculate the Bayesian p-value
    d[i,j,k]<-  abs(Y[i,j,k] - p.eff[i,j,k])
    dnew[i,j,k]<- abs(Ynew[i,j,k] - p.eff[i,j,k])
    d2[i,j,k]<- pow(d[i,j,k],2)
    dnew2[i,j,k]<- pow(dnew[i,j,k],2)
    }
    dsum[i,k]<- sum(d2[i,1:nrep,k])
    dnewsum[i,k]<- sum(dnew2[i,1:nrep,k])
    }
    }
    # Derived quantities
    # discrepancy measure
    p.fit<-sum(dsum[1:nsite,1:nspec])
    p.fitnew<-sum(dnewsum[1:nsite,1:nspec])

    for(k in 1:(nspec)){
    occ.fs[k] <- sum(z[,k])         # Number of occupied sites
    }
    for (i in 1:nsite) {
       Nsite[i] <- sum(z[i,])   # Number species at each site
    }
    }
    ",fill = TRUE)
sink()

# Initial values
nsite = dim(Y)[1]
nrep = dim(Y)[2]
nspec = dim(Y)[3]
```

```
zst <- tmp  # Observed occurrence as starting values for z
inits <- function() list(z = zst,
                         betalpsi0 = rnorm(n = nspec),
                         betalpsi1 = rnorm(n = nspec),
                            ....
                         betalp9 = rnorm(n = nspec),
                         betalp0 = rnorm(n = nspec),
                         p0 = runif(1,0,1),
                         psi0 = runif(1,0,1),
                         rho = runif(1, 0,1),
                         sigmap = runif(1,0,1.5),
                         sigmapsi =  runif(1,0,1.5),
                         delta = rnorm(n = nspec))

# Parameters monitored
params <- c("Nsite", "occ.fs", "betalpsi0", "betalpsi1", "betalpsi2",
"betalpsi3", "betalpsi4", "betalpsi5", "betalpsi6","betalpsi7",
"betalpsi8","betalp0", "betalp1", "betalp2", "betalp3", "betalp4",
"betalp5", "betalp6", "betalp7", "betalp8", "betalp9",
"mu.betalpsi1", "mu.betalpsi2","mu.betalpsi3", "mu.betalpsi4",
"mu.betalpsi5", "mu.betalpsi6","mu.betalpsi7", "mu.betalpsi8",
"mu.lp", "mu.betalp1", "mu.betalp2", "mu.betalp3", "mu.betalp4",
"mu.betalp5", "mu.betalp6", "mu.betalp7", "mu.betalp8",
"mu.betalp9", "rho", "delta", "p.fit", "p.fitnew")

# MCMC settings
ni <- 70000
nt <- 50
nb <- 20000
nc <- 3
```

A7.3 Summary of values for community hyperparameters

Table A7.1 gives the posterior estimates for the community level mean response to predictors of occupancy and detection (shown in Figure 7.6 in the main text).

Table A7.1. Posterior summaries for the mean (hyperparameter) for community-level effects of the covariates of interest

	Mean (95% Credibility Interval)
Occupancy	
Household size	0.24 (0.01 – 0.48)
Maximum LST	-0.06 (-0.45 – 0.29)
Mean NDVI	0.00 (-0.52 – 0.53)
Population density	0.05 (-0.34 – 0.52)
Owns a cow	0.04 (-0.45 – 0.59)
Grass area	-0.17 (-0.46 – 0.10)
SEP	0.00 (-0.29 – 0.32)
Surface water	-0.04 (-0.55 – 0.42)
Detection	
Age	0.05 (-0.43 – 0.55)
HIV infection	-0.46 (-1.20 – 0.36)
Male gender	0.11 (-0.61 – 0.83)
Milk everyday	-0.16 (-0.61 – 0.21)
SEP	-0.21 (-0.58 – 0.15)
Household size	-0.05 (-0.27 – 0.16)
Latrine	-0.20 (-0.81 – 0.40)
Vitamin A deficiency	0.21 (-0.61 – 1.10)
Iron deficiency	0.13 (-0.26 – 0.47)

A7.4 Outputs from the non-spatial model

In order to explore the effect of accounting for spatial structure on model inference, we repeated the modelling procedure described in the main text, but without the spatial autocovariate term.

Overall inference was the same, but there was much stronger evidence of an effect of maximum LST on *S. mansoni* infection (in the spatial model, the 95% CI overlapped zero whilst it did not in the non-spatial model) (Figure A7.1). A similar effect was observed for the effect of NDVI on *Taenia* and grass area on *Trichuris*, both of which were non-significant in the spatial model. The positive effect of the use of surface water also became significant in the case of *Trichuris* and *Schistosoma*.

Despite these effects at the species-level, there were no obvious differences in the community-level estimates (data not shown).

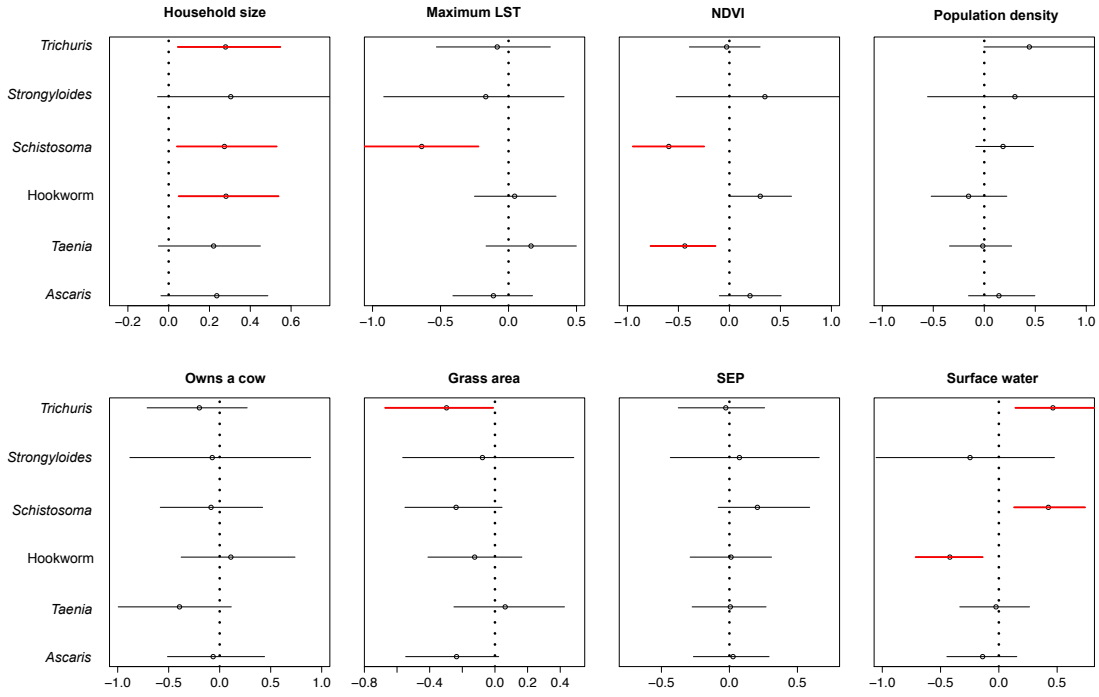


Figure A7.1. Species-level effects of covariates on the probability of occupancy from the non-spatial model. Points show posterior mean, horizontal bars show range of 95% credibility intervals (red lines do not include zero).

Bibliography

Aagaard-Hansen, J. & Chaignat, C.L., 2010. *Neglected Tropical Diseases: Equity and Social Determinants*, World Health Organization : Geneva.

Abdi, H., Williams, L.J. & Valentin, D., 2013. Multiple factor analysis: Principal component analysis for multi-table and multi-block data sets. *Computational Statistics*, 5, pp.149–179.

Abdolrasouli, A., McMillan, A. & Ackers, J.P., 2009. Sexual transmission of intestinal parasites in men who have sex with men. *Sexual Health*, 6(3), pp.185–194.

Actor, J.K. et al., 1993. Helminth infection results in decreased virus-specific CD8+ cytotoxic T-cell and Th1 cytokine responses as well as delayed virus clearance. *Proceedings of the National Academy of Sciences of the United States of America*, 90(3), pp.948–952.

Adazu, K. et al., 2005. Health and demographic surveillance in rural western Kenya: a platform for evaluating interventions to reduce morbidity and mortality from infectious diseases. *American Journal of Tropical Medicine and Hygiene*, 73(6), pp.1151–1158.

Adegnika, A.A. et al., 2010. Epidemiology of parasitic co-infections during pregnancy in Lambarene, Gabon. *Tropical Medicine and International Health*, 15(10), pp.1204–1209.

Ahmed, M.M., Jabbar, M. & Ehui, S., 2000. Household-level economic and nutritional impacts of market-oriented dairy production in the Ethiopian highlands. *Food & Nutrition Bulletin*, 21(4), pp.460–465.

Allan, J.C. et al., 1996. Field trial of the coproantigen-based diagnosis of Taenia solium taeniasis by enzyme-linked immunosorbent assay. *American Journal of Tropical Medicine and Hygiene*, 54(4), pp.352–356.

Allan, J.C. et al., 1990. Immunodiagnosis of Taeniasis by Coproantigen Detection. *Parasitology*, 101, pp.473–477.

Allen, A.V. & Ridley, D.S., 1970. Further observations on the formol-ether concentration technique for faecal parasites. *Journal of Clinical Pathology*, 23(6), pp.545–546.

Allotey, P. et al., 2008. Efficacious, effective, and embedded interventions: implementation research in infectious disease control. *BMC Public Health*, 8, p.343.

Alsan, M.M. et al., 2011. Poverty, global health, and infectious disease: lessons from Haiti and Rwanda. *Infectious Disease Clinics of North America*, 25(3), pp.611.

Alvarez-Dardet, C., 2000. Measuring inequalities in health in Africa. *Journal of Epidemiology and Community Health*, 54(5), p.321.

Amornkul, P.N. et al., 2009. HIV prevalence and associated risk factors among individuals aged 13-34 years in Rural Western Kenya. *PLoS ONE*, 4(7), pp.e6470.

Anderson, R., Truscott, J. & Hollingsworth, T.D., 2014. The coverage and frequency of mass drug administration required to eliminate persistent transmission of soil-transmitted helminths. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 369(1645).

Anderson, R.M. & May, R.M., 1992. *Infectious Diseases of Humans*, Oxford University Press.

Anderson, R.M. & Schad, G.A., 1985. Hookworm burdens and faecal egg counts: an analysis of the biological basis of variation. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 79(6), pp.812–825.

Appleton, C.C. et al., 2009. Geohelminth infection and re-infection after chemotherapy among slum-dwelling children in Durban, South Africa. *Annals of Tropical Medicine and Parasitology*, 103(3), pp.249–261.

Asaolu, S.O. & Ofoezie, I.E., 2003. The role of health education and sanitation in the control of helminth infections. *Acta tropica*, 86(2-3), pp.283–294.

Ashford, D.A., Craig, P.S. & Oppenheimer, S.J., 1993. Polyparasitism on the Kenya Coast .2. Spatial Heterogeneity in Parasite Distributions. *Annals of Tropical Medicine and Parasitology*, 87(3), pp.283–293.

Assefa, S. et al., 2009. Intestinal parasitic infections in relation to HIV/AIDS status, diarrhea and CD4 T-cell count. *BMC Infectious Diseases*, 9, p.155.

Auchincloss, A.H. et al., 2012. A review of spatial methods in epidemiology, 2000-2010. *Annual Review of Public Health*, 33, pp.107–122.

Audibert, M., 2010. Endemic Diseases and Agricultural Productivity: Challenges and Policy Response. *Journal of African Economies*, 19(3), pp.110–165.

Augustin, N.H., Mugglestone, M.A. & Buckland, S.T., 1996. An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology*, 33(2), pp.339–347.

Awusabo-Asare, K. & Annim, S.K., 2008. Wealth status and risky sexual behaviour in Ghana and Kenya. *Applied Health Economics and Health Policy*, 6(1), pp.27–39.

Ayikukwei, R. et al., 2008. HIV/AIDS and cultural practices in western Kenya: the impact of sexual cleansing rituals on sexual behaviours. *Culture, Health & Sexuality*, 10(6), pp.587–599.

Babyak, M.A., 2004. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine*, 66(3), pp.411–421.

Bae, J.K. et al., 2013. Risk factors associated with Lawsonia intracellularis in English pig farms. *The Veterinary Journal*, 197(3), pp.707–711.

Barry, M.M., 2007. The tail end of guinea worm - global eradication without a drug or a vaccine. *New England Journal of Medicine*, 356(25), pp.2561–2564.

Barter, D.M. et al., 2012. Tuberculosis and poverty: the contribution of patient costs in sub-Saharan Africa--a systematic review. *BMC Public Health*, 12(1), p.980.

Basu, A.M. & Stephenson, R., 2005. Low levels of maternal education and the proximate determinants of childhood mortality: a little learning is not a dangerous thing. *Social Science & Medicine*, 60(9), pp.2011–2023.

Bates, D. et al., 2014. lme4: Linear mixed-effects models using Eigen and S4. *R package version 1.1-7, http://CRAN.R-project.org/package=lme4.*

Bates, I., Fenton, C., Gruber, J., Lalloo, D., Lara, A.M., et al., 2004a. Vulnerability to malaria, tuberculosis, and HIV/AIDS infection and disease. Part II: Determinants operating at

environmental and institutional level. *The Lancet Infectious Diseases*, 4(6), pp.368–375.

Bates, I., Fenton, C., Gruber, J., Lalloo, D., Medina Lara, A., et al., 2004b. Vulnerability to malaria, tuberculosis, and HIV/AIDS infection and disease. Part 1: determinants operating at individual and household level. *The Lancet Infectious Diseases*, 4(5), pp.267–277.

Batwala, V., Magnussen, P. & Nuwaha, F., 2010. Are rapid diagnostic tests more accurate in diagnosis of plasmodium falciparum malaria compared to microscopy at rural health centres? *Malaria Journal*, 9.

Becker, S.L. et al., 2011. Diagnosis, clinical features, and self-reported morbidity of Strongyloides stercoralis and hookworm infection in a Co-endemic setting. *PLoS Neglected Tropical Diseases*, 5(8), p.e1292.

Bell, R., Taylor, S. & Marmot, M., 2010. Global health governance: commission on social determinants of health and the imperative for change. *Journal of Law, Medicine & Ethics*, 38(3), pp.470–485.

Benschop, J. et al., 2010. Bayesian zero-inflated predictive modelling of herd-level Salmonella prevalence for risk-based surveillance. *Zoonoses and Public Health*, 57 Suppl 1, pp.60–70.

Bentwich, Z. et al., 2000. Concurrent infections and HIV pathogenesis. *AIDS*, 14(14), pp.2071–2081.

Bentwich, Z. et al., 1998. Immune activation in the context of HIV infection. *Clinical and Experimental Immunology*, 111(1), pp.1–2.

Berman, P., Kendall, C. & Bhattacharyya, K., 1994. The household production of health: Integrating social science perspectives on micro-level health determinants. *Social Science & Medicine*, 38(2), pp.205–215.

Bethony, J. et al., 2002. Additive host genetic factors influence fecal egg excretion rates during Schistosoma mansoni infection in a rural area in Brazil. *American Journal of Tropical Medicine and Hygiene*, 67(4), pp.336–343.

Bethony, J. et al., 2001. Exposure to Schistosoma mansoni infection in a rural area in Brazil. II: household risk factors. *Tropical Medicine and International Health*, 6(2), pp.136–145.

Bethony, J. et al., 2006. Soil-transmitted helminth infections: ascariasis, trichuriasis, and hookworm. *The Lancet*, 367(9521), pp.1521–1532.

Bégaud, E. et al., 2006. Reduced CD4 T cell activation and in vitro susceptibility to HIV-1 infection in exposed uninfected Central Africans. *Retrovirology*, 3, p.35.

Bisanzio, D. et al., 2014. Cross-sectional study of the burden of vector-borne and soil-transmitted polyparasitism in rural communities of Coast Province, Kenya. *PLoS Neglected Tropical Diseases*, 8(7), p.e2992.

Bisoffi, Z. et al., 2013. Strongyloides stercoralis: a plea for action. *PLoS Neglected Tropical Diseases*, 7(5), p.e2214.

Bithell, J.F., 1990. An application of density estimation to geographical epidemiology. *Statistics in Medicine*, 9(6), pp.691–701.

Black, R.E. et al., 2010. Global, regional, and national causes of child mortality in 2008: a systematic analysis. *Lancet*, 375(9730), pp.1969–1987.

Black, R.E. et al., 2008. Maternal and child undernutrition: global and regional exposures and health consequences. *The Lancet*, 371(9608), pp.243–260.

Blakely, T.A. & Woodward, A.J., 2000. Ecological effects in multi-level studies. *Journal of Epidemiology and Community Health*, 54(5), pp.367–374.

Bled, F., Nichols, J.D. & Altwegg, R., 2013. Dynamic occupancy models for analyzing species' range dynamics across large geographic scales. *Ecology and Evolution*, 3(15), pp.4896–4909.

Boaitey, Y.A. et al., 2012. Gastrointestinal and urinary tract pathogenic infections among HIV seropositive patients at the Komfo Anokye Teaching Hospital in Ghana. *BMC Research Notes*, 5, p.454.

Boccia, D. et al., 2011. The association between household socioeconomic position and prevalent tuberculosis in Zambia: a case-control study. *PLoS ONE*, 6(6), p.e20824.

Boccia, D. et al., 2009. Tuberculosis Infection in Zambia: The Association with Relative Wealth. *American Journal of Tropical Medicine and Hygiene*, 80(6), pp.1004–1011.

Boelaert, M. et al., 2010. Socio-economic aspects of neglected diseases: sleeping sickness and visceral leishmaniasis. *Annals of Tropical Medicine and Parasitology*, 104(7), pp.535–542.

Bogoch, I.I. et al., 2006. Differences in microscopic diagnosis of helminths and intestinal protozoa among diagnostic centres. *European Journal of Clinical Microbiology & Infectious Diseases*, 25(5), pp.344–347.

Booth, M., 2006. The role of residential location in apparent helminth and malaria associations. *Trends in Parasitology*, 22(8), pp.359–362.

Booth, M. et al., 2004. Micro-geographical variation in exposure to Schistosoma mansoni and malaria, and exacerbation of splenomegaly in Kenyan school-aged children. *BMC Infectious Diseases*, 4, p.13.

Booth, M. et al., 2003. The influence of sampling effort and the performance of the Kato-Katz technique in diagnosing Schistosoma mansoni and hookworm co-infections in rural Côte d'Ivoire. *Parasitology*, 127, pp.525–531.

Borkow, G. et al., 2001. Helminths, human immunodeficiency virus and tuberculosis. *Scandinavian Journal of Infectious Diseases*, 33(8), pp.568–571.

Bradley, D.J. & May, R.M., 1978. Consequences of helminth aggregation for the dynamics of schistosomiasis. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 72(3), pp.262–273.

Brady, M.A., Hooper, P.J. & Ottesen, E.A., 2006. Projected benefits from integrating NTD programs in sub-Saharan Africa. *Trends in Parasitology*, 22(7), pp.285–291.

Branda, J.A. et al., 2006. A rational approach to the stool ova and parasite examination. *Clinical Infectious Diseases*, 42(7), pp.972–978.

Branscum, A.J. et al., 2005. Effect of diagnostic testing error on intracluster correlation coefficient estimation. *Preventive Veterinary Medicine*, 69(1-2), pp.63–75.

Branscum, A.J., Gardner, I.A. & Johnson, W.O., 2004. Bayesian modeling of animal- and herd-level prevalences. *Preventive Veterinary Medicine*, 66(1-4), pp.101–112.

Braveman, P. & Tarimo, E., 2002. Social inequalities in health within countries: not only an issue for affluent nations. *Social Science & Medicine*, 54(11), pp.1621–1635.

Bronsvoort, M. et al., 2001. Management factors associated with seropositivity to Lawsonia intracellularis in US swine herds. *Journal of Swine Health and Production*, 9(6), pp.285–290.

Brooker, S., 2010. Estimating the global distribution and disease burden of intestinal nematode infections: adding up the numbers--a review. *International Journal for Parasitology*, 40(10), pp.1137–1144.

Brooker, S. et al., 2000. Epidemiology of single and multiple species of helminth infections among school children in Busia District, Kenya. *East African Medical Journal*, 77(3).

Brooker, S., Alexander, N., et al., 2006a. Contrasting patterns in the small-scale heterogeneity of human helminth infections in urban and rural environments in Brazil. *International Journal for Parasitology*, 36(10), pp.1143–1151.

Brooker, S., Bethony, J. & Hotez, P.J., 2004a. Human hookworm infection in the 21st century. *Advances in Parasitology*, 58, pp.197–288.

Brooker, S., Clarke, S., et al., 2004b. Spatial clustering of malaria and associated risk factors during an epidemic in a highland area of western Kenya. *Tropical Medicine and International Health*, 9(7), pp.757–766.

Brooker, S., Clements, A.C.A. & Bundy, D.A.P., 2006b. Global epidemiology, ecology and control of soil-transmitted helminth infections. *Advances in Parasitology*, 62, pp.221–261.

Brooker, S., Clements, A.C.A., Hotez, P.J., et al., 2006c. The co-distribution of Plasmodium falciparum and hookworm among African schoolchildren. *Malaria Journal*, 5, p.99.

Brooker, S., Hotez, P.J. & Bundy, D.A.P., 2008. Hookworm-Related Anaemia among Pregnant Women: A Systematic Review. *PLoS Neglected Tropical Diseases*, 2(9), p.e291.

Brooker, S.J. et al., 2012. Plasmodium-helminth coinfection and its sources of heterogeneity across East Africa. *The Journal of Infectious Diseases*, 205(5), pp.841–852.

Brown, J.H., 1984. On the relationship between abundance and distribution of species. *American Naturalist*, 124(2).

Brown, M. et al., 2006. Helminths and HIV infection: epidemiological observations on immunological hypotheses. *Parasite Immunology*, 28(11), pp.613–623.

Brudney, K. & Dobkin, J., 1991. Resurgent tuberculosis in New York City. Human immunodeficiency virus, homelessness, and the decline of tuberculosis control programs. *The American Review of Respiratory Disease*, 144(4), pp.745–749.

Brunello, M.E.F. et al., 2011. Areas of vulnerability to HIV/TB co-infection in Southeastern Brazil. *Revista de Saúde Pública*, 45(3), pp.556–563.

Brutus, L. et al., 2007. Confirmation of the protective effect of Ascaris lumbricoides on Plasmodium falciparum infection: results of a randomized trial in Madagascar. *American Journal of Tropical Medicine and Hygiene*, 77(6), pp.1091–1095.

Brutus, L. et al., 2006. Parasitic co-infections: does Ascaris lumbricoides protect against Plasmodium falciparum infection? *American Journal of Tropical Medicine and Hygiene*, 75(2), pp.194–198.

Buck, A.A., Anderson, R.I. & MacRae, A.A., 1978. Epidemiology of poly-parasitism. I. Occurrence, frequency and distribution of multiple infections in rural communities in Chad, Peru, Afghanistan, and Zaire. *Tropenmedizin und Parasitologie*, 29(1), pp.61–70.

Bull, T., 2009. Social determinants of health in very poor ruralities. *Global Health Promotion*, 16(1), pp.53–56.

Bundy, D.A.P & Cooper, E.S., 1989. Trichuris and trichuriasis in humans. *Advances in Parasitology*, 28.

Bundy, D.A.P & Medley, G.F., 1992. Immuno-epidemiology of human geohelminthiasis: ecological and immunological determinants of worm burden. *Parasitology*, 104 Suppl, pp.S105–S119.

Bundy, D.A.P., Walson, J.L. & Watkins, K.L., 2013. Worms, wisdom, and wealth: why deworming can make economic sense. *Trends in Parasitology*, 29(3), pp.142–148.

Burnham, K.P. & Anderson, D.R., 1998. *Model Selection and Inference*, Springer.

Burton, A.C. et al., 2012. Hierarchical multi-species modeling of carnivore responses to hunting, habitat and prey in a West African protected area. *PLoS ONE*, 7(5), p.e38007.

Cairncross, S. et al., 1996. The public and domestic domains in the transmission of disease. *Tropical Medicine and International Health*, 1(1), pp.27–34.

Cairncross, S. et al., 2010. Water, sanitation and hygiene for the prevention of diarrhoea. *International Journal of Epidemiology*, 39 Suppl 1, pp.193–205.

Campbell, S.J. et al., 2014. Water, Sanitation, and Hygiene (WASH): a critical component for sustainable soil-transmitted helminth and schistosomiasis control. *PLoS Neglected Tropical Diseases*, 8(4), p.e2651.

Card, C.M. et al., 2009. Decreased immune activation in resistance to HIV-1 infection is associated with an elevated frequency of CD4(+)CD25(+)FOXP3(+) regulatory T cells. *Journal of Infectious Diseases*, 199(9), pp.1318–1322.

Cartwright, C.P., 1999. Utility of multiple-stool-specimen ova and parasite examinations in a high-prevalence setting. *Journal of Clinical Microbiology*, 37(8), pp.2408–2411.

Chachage, M. et al., 2014. Helminth-associated systemic immune activation and HIV co-receptor expression: response to albendazole/praziquantel treatment. *PLoS Neglected Tropical Diseases*, 8(3), p.e2755.

Chaix, B., Merlo, J. & Chauvin, P., 2005. Comparison of a spatial approach with the multilevel approach for investigating place effects on health: the example of healthcare utilisation in France. *Journal of Epidemiology and Community Health*, 59(6), pp.517–526.

Chambers, R., 1998. Vulnerability, coping and policy. *IDS Bulletin*, 20, pp.1–7.

Chan, L., Bundy, D.A.P & Kan, S.P.S, 1994a. Genetic relatedness as a determinant of predisposition to Ascaris lumbricoides and Trichuris trichiura infection. *Parasitology*, 108, pp.77–80.

Chan, L.L., Bundy, D.A.P. & Kan, S.P.S., 1994b. Aggregation and predisposition to Ascaris lumbricoides and Trichuris trichiura at the familial level. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 88(1), pp.46–48.

Chandola, T.T., 2012. Spatial and social determinants of urban health in low-, middle- and high-

income countries. *Annual Review of Public Health*, 126(3), pp.259–261.

Chang, A.Y. et al., 2009. Combining Google Earth and GIS mapping technologies in a dengue surveillance system for developing countries. *International Journal of Health Geographics*, 8, p.49.

Charlton, K.E. & Rose, D., 2002. Prevalence of household food poverty in South Africa: results from a large, nationally representative survey. *Public Health Nutrition*, 5(3), pp.383–389.

Chatterjee, S. et al., 2014. Incidence of active pulmonary tuberculosis in patients with coincident filarial and/or intestinal helminth infections followed longitudinally in South India. *PLoS ONE*, 9(4), pp.e94603.

Chen, J. et al., 2011. Interferon-Gamma Release Assays for the Diagnosis of Active Tuberculosis in HIV-Infected Patients: A Systematic Review and Meta-Analysis. *PLoS ONE*, 6(11).

Chen, Y.-Y. et al., 2014. New integrated strategy emphasizing infection source control to curb Schistosomiasis japonica in a marshland area of Hubei Province, China: findings from an eight-year longitudinal survey. *PLoS ONE*, 9(2), pp.e89779.

Chenine, A.-L. et al., 2008. Acute Schistosoma mansoni infection increases susceptibility to systemic SHIV clade C infection in rhesus macaques after mucosal virus exposure. *PLoS Neglected Tropical Diseases*, 2(7), p.e265.

Chenine, A.-L. et al., 2005. Schistosoma mansoni infection promotes SHIV clade C replication in rhesus macaques. *AIDS*, 19(16), pp.1793–1797.

Chianura, L. et al., 2006. Neurocysticercosis and human immunodeficiency virus infection: A case report. *Journal of Travel Medicine*, 13(6), pp.376–380.

Chipeta, M.G., Ngwira, B. & Kazembe, L.N., 2013. Analysis of Schistosomiasis haematobium infection prevalence and intensity in Chikhwawa, Malawi: an application of a two part model. *PLoS Neglected Tropical Diseases*, 7(3), p.e2131.

Chongsuvivatwong, V. et al., 1996. Predictors for the risk of hookworm infection: experience from endemic villages in southern Thailand. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 90(6), pp.630–633.

Christensen, J. & Gardner, I.A., 2000. Herd-level interpretation of test results for epidemiologic studies of animal diseases. *Preventive Veterinary Medicine*, 45(1-2), pp.83–106.

Christiaensen, L.J. & Subbarao, K., 2005. Towards an Understanding of Household Vulnerability in Rural Kenya. *Journal of African Economies*, 14(4), pp.520–558.

Chukwuocha, U.M. et al., 2014. Determinants of Delay in Seeking Malaria Treatment for Children Under-Five Years in Parts of South Eastern Nigeria. *Journal of Community Health*, 39(6): pp. 1171 - 1178

Chuma, J.M., Thiede, M. & Molyneux, C.S., 2006. Rethinking the economic costs of malaria at the household level: evidence from applying a new analytical framework in rural Kenya. *Malaria Journal*, 5, p.76.

Clements, A.C.A., Deville, M.-A., et al., 2010a. Spatial co-distribution of neglected tropical diseases in the east African great lakes region: revisiting the justification for integrated control. *Tropical*

Medicine and International Health, 15(2), pp.198–207.

Clements, A.C.A., Kur, L.W., et al., 2010b. Targeting trachoma control through risk mapping: the example of Southern Sudan. *PLoS Neglected Tropical Diseases*, 4(8), p.e799.

Clennon, J.A. et al., 2004. Spatial patterns of urinary schistosomiasis infection in a highly endemic area of coastal Kenya. *American Journal of Tropical Medicine and Hygiene*, 70(4), pp.443–448.

Clifford, P., Richardson, S. & Hémon, D., 1989. Assessing the significance of the correlation between two spatial processes. *Biometrics*, 45(1), pp.123–134.

Cohen, J.M., Wilson, M.L. & Aiello, A.E., 2007. Analysis of social epidemiology research on infectious diseases: historical patterns and future opportunities. *Journal of Epidemiology and Community Health*, 61(12), pp.1021–1027.

Conelly, W.T. & Chaiken, M.S., 2000. Intensive farming, agro-diversity, and food security under conditions of extreme population pressure in Western Kenya. *Human Ecology*, 28(1), pp.19–49.

Conteh, L., Engels, T. & Molyneux, D.H., 2010. Socioeconomic aspects of neglected tropical diseases. *Lancet*, 375(9710), pp.239–247.

Conway, D.J. et al., 1995. Household Aggregation of Strongyloides stercoralis Infection in Bangladesh. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 89(3), pp.258–261.

Cook, E.A., 2014. *The epidemiology of zoonoses in slaughterhouse workers in western Kenya*. Unpublished PhD thesis.

Corbett, E.L. et al., 2003. The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. *Archives of Internal Medicine*, 163(9), pp.1009–1021.

Coreil, J., Whiteford, L. & Salazar, D., 2000. The Household Ecology of Disease Transmission: Dengue Fever in the Dominican Republic. In P. J. Brown & M. C. Inhorn, eds. *The Anthropology of Infectious Disease: International Perspectives*. Routledge, p. 483.

Cortinovis, I., Vella, V. & Ndiku, J., 1993. Construction of a socio-economic index to facilitate analysis of health data in developing countries. *Social Science & Medicine*, 36(8), pp.1087–1097.

Costard, S. et al., 2009. Multivariate analysis of management and biosecurity practices in smallholder pig farms in Madagascar. *Preventive Veterinary Medicine*, 92(3), pp.199–209.

Cox, F., 1992. Basic laboratory methods in medical parasitology. *Parasitology Today*, 8(1), pp.35–35.

Cox, F.E., 2001. Concomitant infections, parasites and immune responses. *Parasitology*, 122 Suppl, pp.S23–38.

Criscione, C.D.C. et al., 2010. Landscape genetics reveals focal transmission of a human macroparasite. *PLoS Neglected Tropical Diseases*, 4(4), p.e665.

Crompton, D., 2001. Ascaris and ascariasis. *Advances in Parasitology*, 48, pp.286–375.

CSDH, 2008. *Closing the Gap in a Generation: Health Equity through Action on the Social Determinants of Health*, World Health Organization : Geneva.

Cuadros, D.F. & Abu-Raddad, L.J., 2014. Spatial variability in HIV prevalence declines in several countries in sub-Saharan Africa. *Health & Place*, 28, pp.45–49.

Cuadros, D.F. et al., 2011. Effect of variable transmission rate on the dynamics of HIV in sub-Saharan

Africa. *BMC Infectious Diseases*, 11, pp.216–216.

Cuadros, D.F., Awad, S.F. & Abu-Raddad, L.J., 2013. Mapping HIV clustering: a strategy for identifying populations at high risk of HIV infection in sub-Saharan Africa. *International Journal of Health Geographics*, 12, pp.28–28.

Curtis, S. & Rees Jones, I., 1998. Is There a Place for Geography in the Analysis of Health Inequality? *Sociology of Health & Illness*, 20, pp.645–672.

Czachor, J.S. & Jonas, A.P., 2000. Transmission of Strongyloides stercoralis person to person. *Journal of Travel Medicine*, 7(4), pp.211–212.

Dacombe, R.J. et al., 2007. Time delays between patient and laboratory selectively affect accuracy of helminth diagnosis. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 101(2), pp.140–145.

Dai, Y. et al., 2012. Evaluation of interferon-gamma release assays for the diagnosis of tuberculosis: an updated meta-analysis. *European Journal of Clinical Microbiology & Infectious Diseases*, 31(11), pp.3127–3137.

Dale, M.R.T. & Fortin, M.-J., 2009. Spatial autocorrelation and statistical tests: Some solutions. *Journal of Agricultural, Biological, and Environmental Statistics*, 14(2), pp.188–206.

Dangour, A.D. et al., 2013. Interventions to improve water quality and supply, sanitation and hygiene practices, and their effects on the nutritional status of children. *Cochrane Database of Systematic Reviews*, (8).

Davey-Smith, S., 2003. *Health inequalities: lifecourse approaches*, The Policy Press, Bristol.

Davies, T.M., Hazelton, M.L. & Marshall, J.C., 2011. sparr: Analyzing Spatial Relative Risk Using Fixed and Adaptive Kernel Density Estimation in R. *Journal of Statistical Software*, 39(1).

de Castro, Marcia Caldas & Fisher, M.G., 2012. Is malaria illness among young children a cause or a consequence of low socioeconomic status? evidence from the united Republic of Tanzania. *Malaria Journal*, 11(1), p.161.

de Castro, Marcia Caldas, Sawyer, D.O. & Singer, B.H., 2007. Spatial patterns of malaria in the Amazon: implications for surveillance and targeted interventions. *Health & Place*, 13(2), pp.368–380.

de Oliveira, E.C. et al., 2013. Geographic information systems and logistic regression for high-resolution malaria risk mapping in a rural settlement of the southern Brazilian Amazon. *Malaria Journal*, 12, p.420.

de Silva, N.R. et al., 2003. Soil-transmitted helminth infections: updating the global picture. *Trends in Parasitology*, 19(12), pp.547–551.

de Vlas, S.J. & Gryseels, B., 1992. Underestimation of Schistosoma mansoni prevalences. *Parasitology Today*, 8(8), pp.274–277.

Dean, A.S. et al., 2013. Epidemiology of Brucellosis and Q Fever in Linked Human and Animal Populations in Northern Togo. *PLoS ONE*, 8(8).

Delobel, P. et al., 2004. Unusual form of neurocysticercosis associated with HIV infection. *European Journal of Neurology*, 11(1), pp.55–58.

Deribew, A. et al., 2009. Tuberculosis and HIV co-infection: its impact on quality of life. *Health and Quality of Life Outcomes*, 7.

Dewalt, B.R. & Pelto, P.J., 1985. Microlevel/Macrolevel Linkages: An introduction to the Issues and a Framework for Analysis. In B. R. Dewalt & P. J. Pelto, eds. *Micro and Macro Levels of Analysis in Anthropology.* Westview, pp. 1–21.

DeWan, A.A. & Zipkin, E.F., 2010. An Integrated Sampling and Analysis Approach for Improved Biodiversity Monitoring. *Environmental Management*, 45(5), pp.1223–1230.

Diel, R., Loddenkemper, R. & Nienhaus, A., 2010. Evidence-Based Comparison of Commercial Interferon-gamma Release Assays for Detecting Active TB. *Chest*, 137(4), pp.952–968.

Diez-Roux, A.V., 1998. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *American Journal of Public Health*, 88(2), pp.216–222.

Diez-Roux, A.V., 2000. Multilevel analysis in public health research. *Annual Review of Public Health*. 21, pp. 171 - 192.

Diez-Roux, A.V. & Aiello, A.E., 2005. Multilevel analysis of infectious diseases. *Journal of Infectious Diseases*, 191 Suppl 1(s1), pp.S25–S33.

Dohoo, I.R., Martin, S.W. & Stryhn, H., 2009. *Veterinary Epidemiologic Research* 2nd ed., VER Inc.

Dorazio, R.M. & Royle, J.A., 2005. Estimating size and composition of biological communities by modeling the occurrence of species. *Journal of the American Statistical Association*, 100(470), pp.389–398.

Dorazio, R.M. et al., 2006. Estimating species richness and accumulation by modeling species occurrence and detectability. *Ecology*, 87(4), pp.842–854.

Dorazio, R.M., Gotelli, N.J. & Ellison, A.M., 2011. Modern Methods of Estimating Biodiversity from Presence-Absence Surveys. In G. Venora, O. Grillo, & J. Lopez-Pujol, eds. *Biodiversity Loss in a Changing Planet*. Rijeka, Croatia: InTech, pp. 277–302.

Dorward, A. et al., 2005. *A Guide To Indicators and Methods For Assessing The Contribution of Livestock Keeping to the Livelihoods of The Poor*, Department of Agricultural Sciences, Imperial College, London.

Douek, D., 2007. HIV disease progression: immune activation, microbes, and a leaky gut. *Topics in HIV Medicine*, 15(4), pp.114–117.

Duc, P.P. et al., 2011. Risk factors for Entamoeba histolytica infection in an agricultural community in Hanam province, Vietnam. *Parasites & Vectors*, 4(1), p.102.

Ebrahim, A. et al., 1997. Evaluation of the Kato-Katz thick smear and formol ether sedimentation techniques for quantitative diagnosis of Schistosoma mansoni infection. *American Journal of Tropical Medicine and Hygiene*, 57(6), pp.706–708.

Ekpo, U.F. et al., 2008. Helminthiasis and hygiene conditions of schools in Ikenne, Ogun State, Nigeria. *PLoS Neglected Tropical Diseases*, 2(1), p.e146.

Elias, D. et al., 2006. Are intestinal helminths risk factors for developing active tuberculosis? *Tropical Medicine and International Health*, 11(4), pp.551–558.

Elias, D. et al., 2007. Chronic helminth infections may negatively influence immunity against

tuberculosis and other diseases of public health importance. *Expert Review of Anti-infective Therapy*, 5(3), pp.475–484.

Elias, D. et al., 2001. Effect of deworming on human T cell responses to mycobacterial antigens in helminth-exposed individuals before and after bacille Calmette-Guérin (BCG) vaccination. *Clinical and Experimental Immunology*, 123(2), pp.219–225.

Elith, J. & Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology*, 40(1), p.677.

Elliott, A.M. et al., 2003. Associations between helminth infection and CD4+ T cell count, viral load and cytokine responses in HIV-1-infected Ugandan adults. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 97(1), pp.103–108.

Ellis, M.K. & McManus, D.P., 2009. Familial aggregation of human helminth infection in the Poyang lake area of China with a focus on genetic susceptibility to schistosomiasis japonica and associated markers of disease. *Parasitology*, 136(7), pp.699–712.

Elsafi, S.H. et al., 2013. Comparison of microscopy, rapid immunoassay, and molecular techniques for the detection of Giardia lamblia and Cryptosporidium parvum. *Parasitology Research*, 112(4), pp.1641–1646.

Else, K.J. et al., 1994. Cytokine-mediated regulation of chronic intestinal helminth infection. *The Journal of Experimental Medicine*, 179(1), pp.347–351.

Emerson, P.M. et al., 2004. Role of flies and provision of latrines in trachoma control: cluster-randomised controlled trial. *Lancet*, 363(9415), pp.1093–1098.

Engels, D. & Savioli, L., 2006. Reconsidering the underestimated burden caused by neglected tropical diseases. *Trends in Parasitology*, 22(8), pp.363–366.

Ensink J.H. et al., 2005. High risk of hookworm infection among wastewater farmers in Pakistan. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 99(11); 809-818

Erb, K.J. et al., 2002. Infection with the helminth Nippostrongylus brasiliensis does not interfere with efficient elimination of Mycobacterium bovis BCG from the lungs of mice. *Clinical and Diagnostic Laboratory Immunology*, 9(3), pp.727–730.

Erhardt, J.G. et al., 2004. Combined measurement of ferritin, soluble transferrin receptor, retinol binding protein, and C-reactive protein by an inexpensive, sensitive, and simple sandwich enzyme-linked immunosorbent assay technique. *Journal of Nutrition*, 134(11), pp.3127–3132.

Ericsson, C.D. et al., 2001. Diagnosis of Strongyloides stercoralis infection. *Clinical Infectious Diseases*, 33(7), pp.1040–1047.

Ernst, K.C. et al., 2009. Environmental, socio-demographic and behavioural determinants of malaria risk in the western Kenyan highlands: a case-control study. *Tropical Medicine and International Health*, 14(10), pp.1258–1265.

Ervasti, J. et al., 2012. Students school satisfaction as predictor of teachers sickness absence: a prospective cohort study. *European Journal of Public Health*, 22(2), pp.215–219.

Eshetu, E.B. & Woldesenbet, S.A., 2011. Are there particular social determinants of health for the world's poorest countries? *African Health Sciences*, 11(1), pp.108–115.

Esrey, S.A. et al., 1991. Effects of improved water supply and sanitation on ascariasis, diarrhoea, dracunculiasis, hookworm infection, schistosomiasis, and trachoma. *Bulletin of the World Health Organization*, 69(5), pp.609–621.

Ezenwa, V.O. et al., 2010. Hidden consequences of living in a wormy world: nematode-induced immune suppression facilitates tuberculosis invasion in African buffalo. *The American Naturalist*, 176(5), pp.613–624.

Ezzati, M. et al., 2003. Estimates of global and regional potential health gains from reducing muliple major risk factors. *The Lancet*, 362(9380), pp.271–280.

Fançony, C. et al., 2013. Performance of microscopy and RDTs in the context of a malaria prevalence survey in Angola: a comparison using PCR as the gold standard. *Malaria Journal*, 12(1), p.284.

Farmer, P., 2001. *Infections and Inequalities*, Univ of California Press.

Farmer, P., 1996. Social inequalities and emerging infectious diseases. *Emerging Infectious Diseases*, 2(4), pp.259–269.

Farmer, P.E., 2013. Chronic Infectious Disease and the Future of Health Care Delivery. *New England Journal of Medicine*, 369(25), pp.2424–2436.

Feachem, R.G. et al., 1983. Excreta disposal facilities and intestinal parasitism in urban Africa: preliminary studies in Botswana, Ghana and Zambia. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 77(4), pp.515–521.

Feitosa, G. et al., 2001. High prevalence of giardiasis and stronglyloidiasis among HIV-infected patients in Bahia, Brazil. *The Brazilian Journal of Infectious Diseases*, 5(6), pp.339–344.

Feldacker, C., Ennett, S.T. & Speizer, I., 2011. It's not just who you are but where you live: An exploration of community influences on individual HIV status in rural Malawi. *Social Science & Medicine*, 72(5), pp.717–725.

Fenton, A., 2013. Dances with worms: the ecological and evolutionary impacts of deworming on coinfecting pathogens. *Parasitology*, 140(9), pp.1119–1132.

Fenwick, A., 2006. Waterborne infectious diseases - Could they be consigned to history? *Science*, 313(5790), pp.1077–1081.

Ferrier, S. & Guisan, A., 2006. Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, 43(3), pp.393–404.

Field, P.R. et al., 2002. Evaluation of a novel commercial enzyme-linked immunosorbent assay detecting Coxiella burnetii-specific immunoglobulin G for Q fever prevaccination screening and diagnosis. *Journal of Clinical Microbiology*, 40(9), pp.3526–3529.

Filmer, D., 2005. Fever and its treatment among the more and less poor in sub-Saharan Africa. *Health Policy and Planning*, 20(6), pp.337–346.

Filmer, D. & Pritchett, L.H., 2001. Estimating wealth effects without expenditure data--or tears: an application to educational enrollments in states of India. *Demography*, 38(1), pp.115–132.

Fincham, J.E., Markus, M.B. & Adams, V.J., 2003. Could control of soil-transmitted helminthic infection influence the HIV/AIDS pandemic. *Acta Tropica*, 86(2-3), pp.315–333.

Fine, P., Eames, K. & Heymann, D.L., 2011. "Herd immunity": a rough guide. *Clinical Infectious*

Diseases, 52(7), pp.911–916.

Fiske, I. & Chandler, R., 2011. unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software*, 43(10).

Fleury, A. et al., 2003. Detection of secreted cysticercal antigen: a useful tool in the diagnosis of inflammatory neurocysticercosis. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 97(5), pp.542–546.

Flisser, A., 2006. Where are the tapeworms? *Parasitology International*, 55, pp.S117–S120.

Florey, L.S. et al., 2012. Partnering Parasites: Evidence of Synergism between Heavy Schistosoma haematobium and Plasmodium Species Infections in Kenyan Children. *PLoS Neglected Tropical Diseases*, 6(7), p.e1723.

Fontanet, A.L. et al., 2000. Epidemiology of HIV and Schistosoma mansoni infections among sugar-estate residents in Ethiopia. *Annals of Tropical Medicine and Parasitology*, 94(2), pp.145–155.

Forrester, J.E. et al., 1988. Clustering of Ascaris lumbricoides and Trichuris trichiura infections within households. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 82(2), pp.282–288.

Forrester, J.E. et al., 1990. Predisposition of individuals and families in Mexico to heavy infection with Ascaris lumbricoides and Trichuris trichiura. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 84(2), pp.272–276.

Fotedar, R. et al., 2007. Laboratory diagnostic techniques for Entamoeba species. *Clinical Microbiology Reviews*, 20(3), pp.511.

Fotso, J.-C. & Kuate-Defo, B., 2005. Socioeconomic inequalities in early childhood malnutrition and morbidity: modification of the household-level effects by the community SES. *Health & Place*, 11(3), pp.205–225.

Fox, M.P. et al., 2004. The impact of HIV/AIDS on labour productivity in Kenya. *Tropical Medicine and International Health*, 9(3), pp.318–324.

Frantz, F.G. et al., 2007. The immune response to toxocariasis does not modify susceptibility to Mycobacterium tuberculosis infection in BALB/c mice. *American Journal of Tropical Medicine and Hygiene*, 77(4), pp.691–698.

Fürst, T. et al., 2012. Schistosomiasis, soil-transmitted helminthiasis, and socio-demographic factors influence quality of life of adults in Côte d'Ivoire. *PLoS Neglected Tropical Diseases*, 6(10), p.e1855.

Gakidou, E. et al., 2010. Increased educational attainment and its effect on child mortality in 175 countries between 1970 and 2009: a systematic analysis. *Lancet*, 376(9745), pp.959–974.

Gallagher, M. et al., 2005. The effects of maternal helminth and malaria infections on mother-to-child HIV transmission. *AIDS*, 19(16), pp.1849–1855.

Garbossa, G. et al., 2013. Social and environmental health determinants and their relationship with parasitic diseases in asymptomatic children from a shantytown in Buenos Aires, Argentina. *Pathogens and Global Health*, 107(3), pp.141–152.

Garcia, H.H. et al., 2000. Serum antigen detection in the diagnosis, treatment, and follow-up of

neurocysticercosis patients. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 94(6), pp.673–676.

García, H.H. et al., 2003. Taenia solium cysticercosis. *Lancet*, 362(9383), pp.547–556.

Gaze, S. et al., 2012. Characterising the mucosal and systemic immune responses to experimental human hookworm infection. *PLoS Pathogens*, 8(2), p.e1002520.

Gazzinelli, A. et al., 2012. A research agenda for helminth diseases of humans: social ecology, environmental determinants, and health systems. *PLoS Neglected Tropical Diseases*, 6(4), p.e1603.

Gelman, A. & Hill, J., 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press.

Gelman, A. et al., 2013. *Bayesian Data Analysis, Third Edition*, CRC Press.

Gelman, A., Meng, X.L. & Stern, H., 1996. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6(4), pp.733–760.

Genser, B. et al., 2008. Impact of a city-wide sanitation intervention in a large urban centre on social, environmental and behavioural determinants of childhood diarrhoea: analysis of two cohort studies. *International Journal of Epidemiology*, 37(4), pp.831–840.

George, P.J. et al., 2012. Evidence of microbial translocation associated with perturbations in T cell and antigen-presenting cell homeostasis in hookworm infections. *PLoS Neglected Tropical Diseases*, 6(10), p.e1830.

Gerns, H.L., Sangaré, L.R. & Walson, J.L., 2012. Integration of deworming into HIV care and treatment: a neglected opportunity. *PLoS Neglected Tropical Diseases*, 6(7), p.e1738.

Giacometti, A. et al., 1997. Epidemiologic features of intestinal parasitic infections in Italian mental institutions. *European Journal of Epidemiology*, 13(7), pp.825–830.

Gibson, L.R., Li, B. & Remold, S.K., 2010. Treating cofactors can reverse the expansion of a primary disease epidemic. *BMC Infectious Diseases*, 10, pp.248.

Githeko, A.K. et al., 1992. The reservoir of Plasmodium falciparum malaria in a holoendemic area of western Kenya. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 86(4), pp.355–358.

Githinji, S. et al., 2010. Mosquito nets in a rural area of Western Kenya: ownership, use and quality. *Malaria Journal*, 9, p.250.

Glenngård, A.H. & Maina, T.M., 2007. Reversing the trend of weak policy implementation in the Kenyan health sector?--a study of budget allocation and spending of health resources versus set priorities. *Health Research Policy and Systems*, 5, p.3.

Glinz, D., N'Guessan, N.A., et al., 2010a. High prevalence of Strongyloides stercoralis among school children in rural Côte d'Ivoire. *Journal of Parasitology*, 96(2), pp.431–433.

Glinz, D., Silué, K.D., et al., 2010b. Comparing diagnostic accuracy of Kato-Katz, Koga agar plate, ether-concentration, and FLOTAC for Schistosoma mansoni and soil-transmitted helminths. *PLoS Neglected Tropical Diseases*, 4(7), pp.e754.

Glynn, J.R. et al., 2000. Tuberculosis: associations with HIV and socioeconomic status in rural

Malawi. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 94(5), pp.500–503.

Goodman, D. et al., 2007. A comparison of methods for detecting the eggs of Ascaris, Trichuris, and hookworm in infant stool, and the epidemiology of infection in Zanzibari infants. *American Journal of Tropical Medicine and Hygiene*, 76(4), pp.725–731.

Gopinath, R. et al., 2000. Filarial infections increase susceptibility to human immunodeficiency virus infection in peripheral blood mononuclear cells in vitro. *Journal of Infectious Diseases*, 182(6), pp.1804–1808.

Gore, C. & Custovic, A., 2004. Protective parasites and medicinal microbes? The case for the hygiene hypothesis. *Primary Care Respiratory Journal*, 13(2), pp.68–75.

Greenacre, M. & Blasius, J., 2006. *Multiple Correspondence Analysis and Related Methods*, CRC Press.

Greenland, S., 2001. Ecologic versus individual-level sources of bias in ecologic estimates of contextual health effects. *International Journal of Epidemiology*, 30(6), pp.1343–1350.

Griffiths, E.C. et al., 2011. The nature and consequences of coinfection in humans. *The Journal of Infection*, 63(3), pp.200–206.

Gu, W.D. & Swihart, R.K., 2004. Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. *Biological Conservation*, 116(2), pp.195–203.

Guignard, S. et al., 2000. Prevalence of enteroparasites in a residence for children in the Córdoba Province, Argentina. *European Journal of Epidemiology*, 16(3), pp.287–293.

Guillera-Arroita, G. et al., 2014. Ignoring imperfect detection in biological surveys is dangerous: a response to 'fitting and interpreting occupancy models'. *PLoS ONE*, 9(7), p.e99571.

Guyatt, H.L. et al., 1990. The relationship between the frequency distribution of Ascaris lumbricoides and the prevalence and intensity of infection in human communities. *Parasitology*, 101, pp.139–143.

Guyatt, H.L.H. et al., 2001. Evaluation of efficacy of school-based anthelmintic treatments against anaemia in children in the United Republic of Tanzania. *Bulletin of the World Health Organization*, 79(8), pp.695–703.

Gürtler, R.E. et al., 1998. Household prevalence of seropositivity for Trypanosoma cruzi in three rural villages in northwest Argentina: environmental, demographic, and entomologic associations. *American Journal of Tropical Medicine and Hygiene*, 59(5), pp.741–749.

Gwatkin, D.R., 2003. How well do health programmes reach the poor? *The Lancet*, 361(9357), pp.540–541.

Gyorkos, T.W. et al., 2013. Impact of health education on soil-transmitted helminth infections in schoolchildren of the Peruvian Amazon: a cluster-randomized controlled trial. *PLoS Neglected Tropical Diseases*, 7(9), p.e2397.

Hailemariam, G. et al., 2004. Intestinal parasitic infections in HIV/AIDS and HIV seronegative individuals in a teaching hospital, Ethiopia. *Japanese Journal of Infectious Diseases*, 57, pp.41–43.

Haines, A. et al., 2013. Promoting health and advancing development through improved housing in low-income settings. *Journal of Urban Health*, 90(5), pp.810–831.

Hajizadeh, M. et al., 2014. Socioeconomic inequalities in HIV/AIDS prevalence in sub-Saharan African countries: evidence from the Demographic Health Surveys. *International Journal for Equity in Health*, 13, p.18.

Halpenny, C.M. et al., 2013. Regional, household and individual factors that influence soil transmitted helminth reinfection dynamics in preschool children from rural indigenous Panamá. C. M. Halpenny et al., eds. *PLoS Neglected Tropical Diseases*, 7(2), p.e2070.

Handzel, T. et al., 2003. Geographic distribution of schistosomiasis and soil-transmitted helminths in Western Kenya: Implications for anthelminthic mass treatment. *American Journal of Tropical Medicine and Hygiene*, 69(3), pp.318–323.

Hanski, I., 1998. Metapopulation dynamics. *Nature*, 396(6706), pp.41–49.

Hanson, T., Johnson, W.O. & Gardner, I.A., 2003. Hierarchical models for estimating herd prevalence and test accuracy in the absence of a gold standard. *Journal of Agricultural, Biological, and Environmental Statistics*, 8(2), pp.223–239.

Hargreaves, J.R. et al., 2007. Assessing household wealth in health studies in developing countries: a comparison of participatory wealth ranking and survey techniques from rural South Africa. *Emerging Themes in Epidemiology*, 4.

Harms, G. & Feldmeier, H., 2002. HIV infection and tropical parasitic diseases - deleterious interactions in both directions? *Tropical Medicine and International Health*, 7(6), pp.479–488.

Harrell, F.E., 2006. *Regression Modeling Strategies* 2nd ed., Springer.

Haswell-Elkins, M., Elkins, D. & Anderson, R.M., 1989. The influence of individual, social group and household factors on the distribution of Ascaris lumbricoides within a community and implications for control strategies. *Parasitology*, 98(1), pp.125–134.

Hay, S.I., Snow, R.W. & Rogers, D.J., 1998. Predicting malaria seasons in Kenya using multitemporal meteorological satellite sensor data. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 92(1), pp.12–20.

Hayward, A.C. & Coker, R.J., 2000. Could a tuberculosis epidemic occur in London as it did in New York? *Emerging Infectious Diseases*, 6(1), pp.12–16.

Hazelton, M.L. & Davies, T.M., 2009. Inference based on kernel estimates of the relative risk function in geographical epidemiology. *Biometrical Journal.* 51(1), pp.98–109.

Heckendorn, F. et al., 2002. Species-specific field testing of Entamoeba spp. in an area of high endemicity. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 96(5), pp.521–528.

Heymann, D.L., 2005. Social, behavioural and environmental factors and their impact on infectious disease outbreaks. *Journal of Public Health Policy*, 26(1), pp.133–139.

Hoddinott, J.F., Headey, D.D. & Dereje, M., 2014. Cows, missing milk markets and nutrition in rural Ethiopia. *Ethiopia Strategy Support Program*.

Hosmer, D., Lemeshow, S. & Sturdivant, R., 2013. *Applied Logistic Regression*, John Wiley & Sons.

Hosseinipour, M.C. et al., 2007. HIV and parasitic infection and the effect of treatment among adult outpatients in Malawi. *Journal of Infectious Diseases*, 195(9), pp.1278–1282.

Hotez, P., 2008. Hookworm and poverty. *Annals of the New York Academy of Sciences*, 1136, pp.38–44.

Hotez, P.J., 2005. Hookworm infections. In R. L. Guerrant, D. H. Walker, & P. F. Weller, eds. *Tropical Infectious Diseases*. Elsevier, pp. 1265–1273.

Hotez, P.J., 2009. Mass drug administration and integrated control for the world's high-prevalence neglected tropical diseases. *Clinical Pharmacology and Therapeutics*, 85(6), pp.659–664.

Hotez, P.J., 2013. NTDs V.2.0: "blue marble health" - neglected tropical disease control and elimination in a shifting health policy landscape. *PLoS Neglected Tropical Diseases*, 7(11), pp.e2570.

Hotez, P.J., 2011. The neglected tropical diseases and the neglected infections of poverty: overview of their common features, global disease burden and distribution, new control tools, and prospects for disease elimination. In *The Causes and Impacts of Neglected Tropical and Zoonotic Diseases: Opportunities for Integrated Intervention Strategies: Workshop Summary*. pp. 221–236.

Hotez, P.J. & Pecoul, B., 2010. "Manifesto" for advancing the control and elimination of neglected tropical diseases. *PLoS Neglected Tropical Diseases*, 4(5), pp.e718–e718.

Hotez, P.J. et al., 2007. Control of neglected tropical diseases. *New England Journal of Medicine*, 357(10), pp.1018–1027.

Hotez, P.J., Fenwick, A. & Kjetland, E.F., 2009a. Africa's 32 cents solution for HIV/AIDS. *PLoS Neglected Tropical Diseases*, 3(5), p.e430.

Hotez, P.J., Fenwick, A., Savioli, L., et al., 2009b. Rescuing the bottom billion through control of neglected tropical diseases. *Lancet*, 373(9674), pp.1570–1575.

Hotez, P.J., Molyneux, D.H., Fenwick, A., et al., 2006a. Incorporating a rapid-impact package for neglected tropical diseases with programs for HIV/AIDS, tuberculosis, and malaria. *PLoS Medicine*, 3(5), p.e102.

Hotez, P.J., Molyneux, D.H., Stillwaggon, E., et al., 2006b. Neglected tropical diseases and HIV/AIDS. *Lancet*, 368(9550), pp.1865–1866.

Houdijk, J.G., Jackson, F. & Kyriazakis, I., 2009. Nutritional sensitivity of resistance to Trichostrongylus colubriformis in lactating ewes. *Veterinary Parasitology*, 160(3), pp.258–266.

Houweling, T.A., Kunst, A.E. & Mackenbach, J.P., 2003. Measuring health inequality among children in developing countries: does the choice of the indicator of economic status matter? *International Journal for Equity in Health*, 2(1), p.8.

Howe, L.D. et al., 2012. Measuring socio-economic position for epidemiological studies in low- and middle-income countries: a methods of measurement in epidemiology paper. *International Journal of Epidemiology*, 41(3), pp.871–886.

Howe, L.D. et al., 2011. Subjective measures of socio-economic position and the wealth index: a comparative analysis. *Health Policy and Planning*, 26(3), pp.223–232.

Howe, L.D., Hargreaves, J.R. & Huttly, S.R.A., 2008. Issues in the construction of wealth indices for the measurement of socio-economic position in low-income countries. *Emerging Themes in Epidemiology*, 5, p.3.

Hox, J., 2010. *Multilevel Analysis: methods and applications*, Routledge.

Hughes, R.G. et al., 2004. Environmental influences on helminthiasis and nutritional status among Pacific schoolchildren. *International Journal of Environmental Health Research*, 14(3), pp.163–177.

Humphries, D. et al., 2011. Epidemiology of hookworm infection in Kintampo North Municipality, Ghana: patterns of malaria coinfection, anemia, and albendazole treatment failure. *The American Journal of Tropical Medicine and Hygiene*, 84(5), pp.792–800.

Hunter, G. et al., 1992. Intestinal parasites in Zambian patients with AIDS. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 86(5), pp.543–545.

Hunter, P.R., Zmirou-Navier, D. & Hartemann, P., 2009. Estimating the impact on health of poor reliability of drinking water interventions in developing countries. *The Science of the Total Environment*, 407(8), pp.2621–2624.

Huston, C. & Petri, W., 1999. Amebiasis: Clinical implications of the recognition of Entamoeba dispar. *Current Infectious Disease Reports*, 1(5), pp.441–447.

Hübner, M.P. et al., 2012. Chronic Helminth Infection Does Not Exacerbate Mycobacterium tuberculosis Infection. *PLoS Neglected Tropical Diseases*, 6(12), p.e1970.

Hürlimann, E. et al., 2014. The epidemiology of polyparasitism and implications for morbidity in two rural communities of Côte d'Ivoire. *Parasites & Vectors*, 7(1), p.81.

IFAD, 1999. *Assessment of Rural Poverty in West and Central Africa*, Rome.

Iknayan, K.J. et al., 2014. Detecting diversity: emerging methods to estimate species diversity. *Trends in Ecology & Evolution*, 29(2), pp.97–106.

Iriemenam, N.C. et al., 2010. Strongyloides stercoralis and the immune response. *Parasitology International*, 59(1), pp.9–14.

IRIN, 2012. *http://www.irinnews.org/report/96379/guinea-sierra-leone-cholera-outbreak-easing*,

IRIN, 2014. *http://www.irinnews.org/report/100426/rising-ebola-cases-strain-health-workers*,

Irmak, H. et al., 2004. Use of the Brucella Igm and IgG flow assays in the serodiagnosis of human brucellosis in an area endemic for brucellosis. *American Journal of Tropical Medicine and Hygiene*, 70(6), pp.688–694.

Janovy, J., Jr, 2002. Concurrent infections and the community ecology of helminth parasites. *Journal of Parasitology*, 88(3), pp.440–445.

Jayashi, C.M. et al., 2012. Seroprevalence and risk factors for Taenia solium cysticercosis in rural pigs of northern Peru. *PLoS Neglected Tropical Diseases*, 6(7), p.e1733.

Jayne, T.S. et al., 2003. Smallholder income and land distribution in Africa: implications for poverty reduction strategies. *Food Policy*, 28(3), pp.253–275.

Jenkins, M.W. & Curtis, V., 2005. Achieving the "good life": Why some people want latrines in rural Benin. *Social Science & Medicine*, 61(11), pp.2446–2459.

Jenkins, M.W. & Scott, B., 2007. Behavioral indicators of household decision-making and demand for sanitation and potential gains from social marketing in Ghana. *Social Science & Medicine*, 64(12), pp.2427–2442.

Jensen, A.V. et al., 2013. The Prevalence of Latent Mycobacterium tuberculosis Infection Based on an Interferon-gamma Release Assay: A Cross-Sectional Survey among Urban Adults in Mwanza, Tanzania. *PLoS ONE*, 8(5).

Jensen, P.K. et al., 2002. Domestic transmission routes of pathogens: the problem of in-house contamination of drinking water during storage in developing countries. *Tropical Medicine and International Health*, 7(7), pp.604–609.

Jia, T.-W. et al., 2012. Soil-transmitted helminth reinfection after drug treatment: a systematic review and meta-analysis. *PLoS Neglected Tropical Diseases*, 6(5), p.e1621.

Jordan, D. & McEwen, S.A., 1998. Herd-level test performance based on uncertain estimates of individual test performance, individual true prevalence and herd true prevalence. *Preventive Veterinary Medicine*, 36(3), pp.187–209.

Kabatereine, N.B. et al., 2010. How to (or Not to) Integrate Vertical Programmes for the Control of Major Neglected Tropical Diseases in Sub-Saharan Africa. *PLoS Neglected Tropical Diseases*, 4(6), p.e755.

Kabatereine, N.B. et al., 2007. Impact of a national helminth control programme on infection and morbidity in Ugandan schoolchildren. *Bulletin of the World Health Organization*, 85(2), pp.91–99.

KAIS, 2013. *Kenya AIDS indicator survey 2012: preliminary report*, Ministry of Health, Kenya.

Kalinkovich, A. et al., 1998. Decreased CD4 and increased CD8 counts with T cell activation is associated with chronic helminth infection. *Clinical and Experimental Immunology*, 114(3), pp.414–421.

Kamadjeu, R., 2009. Tracking the polio virus down the Congo River: a case study on the use of Google Earth in public health planning and mapping. *International Journal of Health Geographics*, 8, p.4.

Kar, K. & Chambers, R., 2008. *Handbook on community-led total sanitation*, Institute of Development Studies.

Karanja, D.M. et al., 1997. Studies on schistosomiasis in western Kenya: I. Evidence for immune-facilitated excretion of schistosome eggs from patients with Schistosoma mansoni and human immunodeficiency virus coinfections. *American Journal of Tropical Medicine and Hygiene*, 56(5), pp.515–521.

Karp, C.L. & Auwaerter, P.G., 2007a. Coinfection with HIV and tropical infectious diseases. I. Protozoal pathogens. *Clinical Infectious Diseases,* 45(9), pp.1208–1213.

Karp, C.L. & Auwaerter, P.G., 2007b. Coinfection with HIV and tropical infectious diseases. II. Helminthic, fungal, bacterial, and viral pathogens. *Clinical Infectious Diseases,* 45(9), pp.1214–1220.

Katz, N., Chaves, A. & Pellegrino, J., 1972. A simple device for quantitative stool thick-smear

technique in Schistosomiasis mansoni. *Revista do Instituto de Medicina Tropical de Sao Paulo*, 14(6), pp.397–400.

Kayeyi, N., Sandøy, I.F. & Fylkesnes, K., 2009. Effects of neighbourhood-level educational attainment on HIV prevalence among young women in Zambia. *BMC Public Health*, 9, pp.310–310.

Keiser, J. & Utzinger, J., 2008. Efficacy of current drugs against soil-transmitted helminth infections: systematic review and meta-analysis. *JAMA*, 299(16), pp.1937–1948.

Keiser, J. et al., 2002. Polyparasitism with Schistosoma mansoni, geohelminths, and intestinal protozoa in rural Côte d'Ivoire. *Journal of Parasitology*, 88(3), pp.461–466.

Kennedy, E. & Haddad, L., 1994. Are pre-schoolers from female-headed households less malnourished? A comparative analysis of results from Ghana and Kenya. *Journal of Development Studies*, 30(3), pp.680–695.

Kenyon, F. et al., 2009. The role of targeted selective treatments in the development of refugia-based approaches to the control of gastrointestinal nematodes of small ruminants. *Veterinary Parasitology*, 164(1), pp.3–11.

Kery, M., 2008. Estimating abundance from bird counts: Binomial mixture models uncover complex covariate relationships. *Auk*, 125(2), pp.336–345.

Kery, M. & Royle, J.A., 2008. Inference About Species Richness and Community Structure Using Species-Specific Occupancy Models in the National Swiss Breeding Bird Survey MHB. In D. L. Thomson & E. G. Cooch, eds. *Modeling Demographic Processes In Marked Populations*. Environmental and Ecological Statistics, Vol. 3. Springer.

Kery, M. & Schaub, M., 2012. *Bayesian Population Analysis Using WinBUGS*, Academic Press.

Kery, M., Gardner, B. & Monnerat, C., 2010. Predicting species distributions from checklist data using site-occupancy models. *Journal of Biogeography*, 37(10), pp.1851–1862.

Kery, M., Guillera-Arroita, G. & Lahoz-Monfort, J.J., 2013. Analysing and mapping species range dynamics using occupancy models M. Patten, ed. *Journal of Biogeography*, 40(8), pp.1463–1474.

Khamadi, S.A. et al., 2008. HIV type 1 genetic diversity in Moyale, Mandera, and Turkana based on env-C2-V3 sequences. *Aids Research and Human Retroviruses*, 24(12), pp.1561–1564.

Khamadi, S.A. et al., 2005. HIV type 1 subtypes in circulation in northern Kenya. *Aids Research and Human Retroviruses*, 21(9), pp.810–814.

Khan, M.U., 1982. Interruption of shigellosis by hand washing. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 76(2), pp.164–168.

Khandker, S., Barnes, D. & Hussain, S., 2010. *Energy Poverty in Rural and Urban India: Are the Energy Poor Also Income Poor?* World Bank.

Khieu, V. et al., 2014. High Prevalence and Spatial Distribution of Strongyloides stercoralis in Rural Cambodia. *PLoS Neglected Tropical Diseases*, 8(6).

KHSSP, 2012. *Health sector strategic and investment plan. July 2013-June 2017*, Government of Kenya.

Kightlinger, L.K.L., Seed, J.R.J. & Kightlinger, M.B.M., 1998. Ascaris lumbricoides intensity in relation to environmental, socioeconomic, and behavioral determinants of exposure to infection in children from southeast Madagascar. *Parasitology Research*, 84(3), pp.480–484.

Kikafunda, J. & Namusoke, H., 2006. Nutritional status of HIV/AIDS orphaned children in households headed by the elderly in Rakai district, Southwestern, Uganda. *AJFAND*, 6(1), pp.1–18.

Kilian, A. et al., 2000. Reliability of malaria microscopy in epidemiological studies: results of quality control. *Tropical Medicine and International Health*, 5(1), pp.3–8.

Killewo J.Z.J. et al., 1991. Patterns of hookworm and Ascaris infection in Dar es Salaam. *Acta Tropica* 48, pp.247-249.

King, C.H., 2010. Parasites and poverty: the case of schistosomiasis. *Acta tropica*, 113(2), pp.95–104.

Kitron, U., 1998. Landscape ecology and epidemiology of vector-borne diseases: tools for spatial analysis. *Journal of Medical Entomology*, 35(4), pp.435–445.

Kjetland, E.F. et al., 2006. Association between genital schistosomiasis and HIV in rural Zimbabwean women. *AIDS*, 20(4), pp.593–600.

Knopp, S. et al., 2008. Diagnosis of soil-transmitted helminths in the era of preventive chemotherapy: effect of multiple stool sampling and use of different diagnostic techniques. *PLoS Neglected Tropical Diseases*, 2(11), pp.e331.

Knopp, S. et al., 2013. From morbidity control to transmission control: time to change tactics against helminths on Unguja Island, Zanzibar. *Acta Tropica*, 128(2), pp.412–422.

Knopp, S. et al., 2010. Patterns and risk factors of helminthiasis and anemia in a rural and a peri-urban community in Zanzibar, in the context of helminth control programs. *PLoS Neglected Tropical Diseases*, 4(5), pp.e681.

Knowles, S.C.L., 2011. The effect of helminth co-infection on malaria in mice: a meta-analysis. *International Journal for Parasitology*, 41(10), pp.1041–1051.

Knowles, S.C.L. et al., 2013.Stability of within-host parasite communities in a wild mammal system. *Proceedings of the Royal Society B: Biological Sciences.* 280(1762), 20130598.

Kobylinski, K.C. et al., 2014. Rationale for the Coadministration of Albendazole and Ivermectin to Humans for Malaria Parasite Transmission Control. *American Journal of Tropical Medicine and Hygiene,* 91(4), pp. 655 - 662.

Koch, G.G., Freeman, D.H. & Freeman, J.L., 1975. Strategies in the Multivariate Analysis of Data from Complex Surveys. *International Statistical Review*, 43(1), p.59.

Kolaczinski, J.H. et al., 2007. Neglected tropical diseases in Uganda: the prospect and challenge of integrated control. *Trends in Parasitology*, 23(10), pp.485–493.

Kolenikov, S. & Angeles, G., 2004. *The use of discrete data in PCA: Theory, simulations, and applications to socioeconomic indices*, Carolina Population Center, University of North Carolina, Chapel Hill.

Kongnyuy, E.J. et al., 2006. Wealth and sexual behaviour among men in Cameroon. *BMC International Health and Human Rights*, 6, p.11.

Kongs, A. et al., 2001. The unreliability of the Kato-Katz technique limits its usefulness for evaluating S. mansoni infections. *Tropical Medicine and International Health*, 6(3), pp.163–169.

Kongstad, P. & Mønsted, M., 1980. *Family, Labour, and Trade in Western Kenya*, Nordic Africa Institute.

Koning, F.A. et al., 2005. Low-level CD4+ T cell activation is associated with low susceptibility to HIV-1 infection. *Journal of Immunology*, 175(9), pp.6117–6122.

Koroma, J.B. et al., 2010. Geographical distribution of intestinal schistosomiasis and soil-transmitted helminthiasis and preventive chemotherapy strategies in Sierra Leone. *PLoS Neglected Tropical Diseases*, 4(11), p.e891.

Kosgey, I.S. et al., 2008. Small ruminant production in smallholder and pastoral/extensive farming systems in Kenya. *Small Ruminant Research*, 77(1), pp.11–24.

Krauth, S.J. et al., 2012. An In-Depth Analysis of a Piece of Shit: Distribution of Schistosoma mansoni and Hookworm Eggs in Human Stool. *PLoS Neglected Tropical Diseases*, 6(12), pp.e1969–e1969.

Krefis, A.C. et al., 2011. Spatial analysis of land cover determinants of malaria incidence in the Ashanti Region, Ghana. *PLoS ONE*, 6(3), p.e17905.

Kreft, I.G.G. & de Leeuw, J., 1998. *Introducing Multilevel Modeling*, SAGE.

Krieger, N., 1994. Epidemiology and the web of causation: Has anyone seen the spider? *Social Science & Medicine*, 39(7), pp.887–903.

Kristoffersen, A.B. et al., 2013. Understanding sources of sea lice for salmon farms in Chile. *Preventive Veterinary Medicine*, 111(1-2), pp.165–175.

Krolewiecki, A.J. et al., 2013. A Public Health Response against Strongyloides stercoralis: Time to Look at Soil-Transmitted Helminthiasis in Full. *PLoS Neglected Tropical Diseases*, 7(5), pp.e2165.

Kuhn, I., 2007. Incorporating spatial autocorrelation may invert observed patterns. *Diversity and Distributions*, 13, pp.66–69.

Kuhn, I. & Dormann, C.F., 2012. Less than eight (and a half) misconceptions about spatial analysis. *Journal of Biogeography*, pp.995–1003.

Kullberg, M.C. et al., 1992. Infection with Schistosoma mansoni alters Th1/Th2 cytokine responses to a non-parasite antigen. *Journal of Immunology*, 148(10), pp.3264–3270.

Kulldorff, M., 1997. A spatial scan statistic. *Communications in Statistics*, 26(6), pp.1481–1496.

Kunst, H. & Khan, K.S., 2007. New tests for the diagnosis of latent tuberculosis infection. *Annals of Internal Medicine*, 147(9), pp.672–673.

Lachish, S. et al., 2011. Site-occupancy modelling as a novel framework for assessing test sensitivity and estimating wildlife disease prevalence from imperfect diagnostic tests. *Methods in Ecology and Evolution*, 3(2), pp.339–348.

Lambin, E.F. et al., 2010. Pathogenic landscapes: interactions between land, people, disease vectors, and their animal hosts. *International Journal of Health Geographics*, 9(1), p.54.

Larsen, K. & Merlo, J., 2005. Appropriate assessment of neighborhood effects on individual health:

integrating random and fixed effects in multilevel logistic regression. *American Journal of Epidemiology*, 161(1), pp.81–88.

Lawson, L.L.O. et al., 2004. The stool examination reports amoeba cysts: should you treat in the face of over diagnosis and lack of specificity of light microscopy? *Tropical Doctor*, 34(1), pp.28–30.

Lawton, J.H., 1999. Are there general laws in ecology? *Oikos*, 84(2), pp.177–192.

Le Bailly, M., Landolt, M. & Bouchet, F., 2012. First World War German soldier intestinal worms: an original study of a trench latrine in France. *Journal of Parasitology*, 98(6), pp.1273–1275.

Le Hesran, J.-Y. et al., 2004. Severe malaria attack is associated with high prevalence of Ascaris lumbricoides infection among children in rural Senegal. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 98(7), pp.397–399.

Le, S., Josse, J. & Husson, F., 2008. FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software*, 25(1), pp.1–18.

Legesse, M. et al., 2011. Community-based cross-sectional survey of latent tuberculosis infection in Afar pastoralists, Ethiopia, using QuantiFERON-TB Gold In-Tube and tuberculin skin test. *BMC Infectious Diseases*, 11.

Lello, J. et al., 2013. The relative contribution of co-infection to focal infection risk in children. *Proceedings of the Royal Society B*, 280.

Leroy, J.L. & Frongillo, E.A., 2007. Can interventions to promote animal production ameliorate undernutrition? *Journal of Nutrition*, 137(10), pp.2311–2316.

Lescano, A.G. et al., 2009. Taenia solium Cysticercosis Hotspots Surrounding Tapeworm Carriers: Clustering on Human Seroprevalence but Not on Seizures J. M. Bethony, ed. *PLoS Neglected Tropical Diseases*, 3(1), p.e371.

Levecke, B. et al., 2011. A comparison of the sensitivity and fecal egg counts of the McMaster egg counting and Kato-Katz thick smear methods for soil-transmitted helminths. *PLoS Neglected Tropical Diseases*, 5(6), pp.e1201–e1201.

Levy, M.Z. et al., 2007. Targeted screening strategies to detect Trypanosoma cruzi infection in children. *PLoS Neglected Tropical Diseases*, 1(3), p.e103.

Lewis, F.I. & Torgerson, P.R., 2012. A tutorial in estimating the prevalence of disease in humans and animals in the absence of a gold standard diagnostic. *Emerging Themes in Epidemiology*, 9(1), pp.9–9.

Liese, B.H. & Schubert, L., 2009. Official development assistance for health-how neglected are neglected tropical diseases? An analysis of health financing. *International Health*, 1(2), pp.141–147.

Lim, Y.A.L. et al., 2009. Intestinal parasitic infections amongst Orang Asli (indigenous) in Malaysia: Has socioeconomic development alleviated the problem? *Tropical Biomedicine*, 26(2), pp.110–122.

Lin, A. et al., 2013. Household Environmental Conditions Are Associated with Enteropathy and Impaired Growth in Rural Bangladesh. *American Journal of Tropical Medicine and Hygiene*, 89(1), pp.130–137.

Lindo, J.F. et al., 1995. Age-Prevalence and Household Clustering of Strongyloides-Stercoralis Infection in Jamaica. *Parasitology*, 110, pp.97–102.

Lindo, J.F. et al., 1998. Intestinal parasitic infections in human immunodeficiency virus (HIV)-positive and HIV-negative individuals in San Pedro Sula, Honduras. *American Journal of Tropical Medicine and Hygiene*, 58(4), pp.431–435.

Lindsay, S.W. et al., 2003. Changes in house design reduce exposure to malaria mosquitoes. *Tropical Medicine and International Health*, 8(6), pp.512–517.

Linehan, M. et al., 2011. Integrated implementation of programs targeting neglected tropical diseases through preventive chemotherapy: proving the feasibility at national scale. *American Journal of Tropical Medicine and Hygiene*, 84(1), pp.5–14.

Link, W.A. & Sauer, J.R., 1996. Extremes in ecology: Avoiding the misleading effects of sampling variation in summary analyses. *Ecology*, 77(5), pp.1633–1640.

Liu, L. et al., 2012. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *Lancet*, 379(9832), pp.2151–2161.

Logan, J.R., 2012. Making a Place for Space: Spatial Thinking in Social Science. *Annual Review of Sociology*, 38.

Lopez, A.D. & Mathers, C.D., 2006. Measuring the global burden of disease and epidemiological transitions: 2002-2030. *Annals of Tropical Medicine and Parasitology*, 100(5-6), pp.481–499.

Lumley, T., 2004. Analysis of complex survey samples. *Journal of Statistical Software*, 9(1), pp.1–19.

Lumley, T., 2010. *Complex Surveys: A Guide to Analysis Using R.* 1st ed., John Wiley and Sons.

Lustigman, S. et al., 2012. A research agenda for helminth diseases of humans: the problem of helminthiases. *PLoS Neglected Tropical Diseases*, 6(4), p.e1582.

Lwambo, N.J. et al., 1999. Patterns of concurrent hookworm infection and schistosomiasis in schoolchildren in Tanzania. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 93(5), pp.497–502.

Lynch, J. & Kaplan, G., 2000. Socioeconomic Position. In L. Berkman & I. Kawachi, eds. *Social Epidemiology*. Oxford Univ Press.

Mableson, H.E. et al., 2014. Neglected zoonotic diseases-the long and winding road to advocacy. *PLoS Neglected Tropical Diseases*, 8(6), p.e2800.

MacKenzie, D.I. & Bailey, L.L., 2004. Assessing the fit of site-occupancy models. *Journal of Agricultural, Biological, and Environmental Statistics*, 9(3), pp.300–318.

MacKenzie, D.I. & Nichols, J.D., 2004. Occupancy as a surrogate for abundance estimation. *Animal Biodiversity and Conservation*, 21(1), pp.461–467.

MacKenzie, D.I. et al., 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83(8), pp.1–8.

MacKenzie, D.I. et al., 2006. *Occupancy estimation and modeling: Inferring Patterns and Dynamics of Species Occurrence*, Elsevier.

MacNeill, M.A. et al., 2008. Accounting for detectability in reef-fish biodiversity estimates. *Marine Ecology Progress Series*, 367, pp.249–260.

Magalhães, R.J.S. et al., 2011. The applications of model-based geostatistics in helminth epidemiology and control. *Advances in Parasitology*, 74, pp.267–296.

Maggi, E. et al., 1994. Ability of HIV to promote a TH1 to TH0 shift and to replicate preferentially in TH2 and TH0 cells. *Science*, 265(5169), pp.244–248.

Mandell, G.L., Bennett, J.E. & Dolin, R., 2010. *Mandell, Douglas, and Bennett's principles and practice of infectious diseases*, Churchill Livingstone.

Manderson, L. et al., 2009. Social research on neglected diseases of poverty: continuing and emerging themes. *PLoS Neglected Tropical Diseases*, 3(2), p.e332.

Manne, J. & Maciag, K., 2011. Neglected tropical diseases and the HIV prevention agenda. *Lancet*, 378(9809), pp.2069–2070.

Mansfield, L.S. et al., 1996. Strongyloides stercoralis: Maintenance of exceedingly chronic infections. *American Journal of Tropical Medicine and Hygiene*, 55(6), pp.617–624.

Marchal, B., Cavalli, A. & Kegels, G., 2009. Global health actors claim to support health system strengthening: is this reality or rhetoric? *PLoS Medicine*, 6(4), pp.e1000059.

Marchi Blatt, J. & Cantos, G.A., 2003. Evaluation of techniques for the diagnosis of Strongyloides stercoralis in human immunodeficiency virus (HIV) positive and HIV negative individuals in the city of Itajaí, Brazil. *Brazilian Journal of Infectious Diseases*, 7(6), pp.402–408.

Marmot, M., 2005. Social determinants of health inequalities. *Lancet*, 365(9464), pp.1099–1104.

Marmot, M. & Wilkinson, R., 2005. *Social Determinants of Health*, Oxford University Press.

Marti, H. & Koella, J.C., 1993. Multiple stool examinations for ova and parasites and rate of false-negative results. *Journal of Clinical Microbiology*, 31(11), pp.3044–3045.

Martin, T.G. et al., 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology Letters*, 8(11), pp.1235–1246.

Masset, E. et al., 2012. Effectiveness of agricultural interventions that aim to improve nutritional status of children: systematic review. *BMJ (Clinical research ed.)*, 344, p.d8222.

Mata, L., 1982. Sociocultural factors in the control and prevention of parasitic diseases. *Reviews of Infectious Diseases*, 4(4), pp.871–879.

Mathers, C.D. & Loncar, D., 2006. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Medicine*, 3(11), p.e442.

Matthys, B. et al., 2011. Prevalence and risk factors of helminths and intestinal protozoa infections among children from primary schools in western Tajikistan. *Parasites & Vectors*, 4(1), p.195.

Mbabazi, P.S. et al., 2011. Examining the relationship between urogenital schistosomiasis and HIV infection. *PLoS Neglected Tropical Diseases*, 5(12), p.e1396.

McCarthy, J.S. et al., 2012. A research agenda for helminth diseases of humans: diagnostics for control and elimination programmes. *PLoS Neglected Tropical Diseases*, 6(4), p.e1601.

McDermott, J.J. & Schukken, Y.H., 1994. A review of methods used to adjust for cluster effects in explanatory epidemiological studies of animal populations. *Preventive Veterinary Medicine*, 18(3), pp.155–173.

McLaren, L. & Hawe, P., 2005. Ecological perspectives in health research. *Journal of Epidemiology*

and Community Health, 59(1), pp.6–14.

McManus, D.P. et al., 2014. Health education and the control of intestinal worm infections in China: a new vision. *Parasites & Vectors*, 7(1), p.344.

McMichael, A.J. & Beaglehole, R., 2000. The changing global context of public health. *The Lancet*, 356(9228), pp.495–499.

McV Messam, L.L. et al., 2008. Frequentist and Bayesian approaches to prevalence estimation using examples from Johne's disease. *Animal Health Research Reviews*, 9(1), pp.1–23.

Mecaskey, J.W., Ngirwamungu, E. & Kilima, P.M., 2003. Integration of trachoma control into primary health care: the Tanzanian experience. *American Journal of Tropical Medicine and Hygiene*, 69(5 Suppl), pp.29–32.

Medeiros, Z. et al., 2012. Using kernel density estimates to investigate lymphatic filariasis in northeast Brazil. *Pathogens and Global Health*, 106(2), pp.113–117.

Mehio-Sibai, A. et al., 2005. A Positive or a Negative Confounding Variable? A Simple Teaching Aid for Clinicians and Students. *Annals of Epidemiology*, 15(6), pp.421–423.

Menzies, D., Pai, M. & Comstock, G., 2007. Meta-analysis: New tests for the diagnosis of latent tuberculosis infection: Areas of uncertainty and recommendations for research. *Annals of Internal Medicine*, 146(5), pp.340–354.

Merlo, J., 2003. Multilevel analytical approaches in social epidemiology: measures of health variation compared with traditional measures of association. *Journal of Epidemiology and Community Health*, 57(8), pp.550–552.

Merlo, J. & Wagner, P., 2013. The tyranny of the averages and the indiscriminate use of risk factors in public health: a call for revolution. *European Journal of Epidemiology*, 28(1), p.148.

Merlo, J. et al., 2006. A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *Journal of Epidemiology and Community Health*, 60(4), pp.290–297.

Merlo, J. et al., 2012. Bringing the individual back to small-area variation studies: a multilevel analysis of all-cause mortality in Andalusia, Spain. *Social Science & Medicine (1982)*, 75(8), pp.1477–1487.

Merlo, J., Chaix, B., Yang, M., Lynch, J. & Rastam, L., 2005a. A brief conceptual tutorial of multilevel analysis in social epidemiology: linking the statistical concept of clustering to the idea of contextual phenomenon. *Journal of Epidemiology and Community Health*, 59(6), pp.443–449.

Merlo, J., Chaix, B., Yang, M., Lynch, J. & Råstam, L., 2005b. A brief conceptual tutorial on multilevel analysis in social epidemiology: interpreting neighbourhood differences and the effect of neighbourhood characteristics on individual health. *Journal of Epidemiology and Community Health*, 59(12), pp.1022–1028.

Merlo, J., Yang, M., et al., 2005c. A brief conceptual tutorial on multilevel analysis in social epidemiology: investigating contextual phenomena in different groups of people. *Journal of Epidemiology and Community Health*, 59(9), pp.729–736.

Messer, L.C. et al., 2008. Socioeconomic domains and associations with preterm birth. *Social Science*

& Medicine, 67(8), pp.1247–1257.

Metcalf, C.J.E. et al., 2014. Transport networks and inequities in vaccination: remoteness shapes measles vaccine coverage and prospects for elimination across Africa. *Epidemiology and Infection*, pp.1–10.

Metcalfe, J.Z. et al., 2011. Interferon-gamma Release Assays for Active Pulmonary Tuberculosis Diagnosis in Adults in Low- and Middle-Income Countries: Systematic Review and Meta-analysis. *Journal of Infectious Diseases*, 204, pp.S1120–S1129.

Midzi, N. et al., 2008. The burden of polyparasitism among primary schoolchildren in rural and farming areas in Zimbabwe. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 102(10), pp.1039–1045.

Mirdha, B.R., 2009. Human strongyloidiasis: often brushed under the carpet. *Tropical Gastroenterology*, 30(1), pp.1–4.

Mizanbayeva, S. et al., 2009. The evaluation of a user-friendly lateral flow assay for the serodiagnosis of human brucellosis in Kazakhstan. *Diagnostic Microbiology and Infectious Disease*, 65(1), pp.14–20.

Mkhize-Kwitshana, Z.L. et al., 2011. The influence of different helminth infection phenotypes on immune responses against HIV in co-infected adults in South Africa. *BMC Infectious Diseases*, 11, p.273.

Mmbando, B.P. et al., 2011. Spatial variation and socio-economic determinants of Plasmodium falciparum infection in northeastern Tanzania. *Malaria Journal*, 10, p.145.

Modjarrad, K. & Vermund, S.H., 2010. Effect of treating co-infections on HIV-1 viral load: a systematic review. *The Lancet Infectious Diseases*, 10(7), pp.455–463.

Modjarrad, K. et al., 2005. Treatment of Intestinal Helminths Does Not Reduce Plasma Concentrations of HIV‐1 RNA in Coinfected Zambian Adults. *Journal of Infectious Diseases*, 192(7), pp.1277–1283.

Molyneux, D.H., 2008. Combating the “other diseases” of MDG 6: changing the paradigm to achieve equity and poverty reduction? *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 102(6), pp.509–519.

Molyneux, D.H. et al., 2009. Neglected tropical diseases and the Global Fund. *Lancet*, 373(9660), pp.296–297.

Molyneux, D.H., Hotez, P.J. & Fenwick, A., 2005. “Rapid-impact interventions”: how a policy of integrated control for Africa's neglected tropical diseases could benefit the poor. *PLoS Medicine*, 2(11), p.e336.

Montes, M., Sawhney, C. & Barros, N., 2010. Strongyloides stercoralis: there but not seen. *Current Opinion in Infectious Diseases*, 23(5), pp.500–504.

Montgomery, M.A., Desai, M.M. & Elimelech, M., 2010. Assessment of latrine use and quality and association with risk of trachoma in rural Tanzania. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 104(4), pp.283–289.

Morales, M.A.G. et al., 1995. Opportunistic and non-opportunistic parasites in HIV-positive and

negative patients with diarrhoea in Tanzania. *Tropical Medicine and Parasitology*, 46(2), pp.109–114.

Mordecai, R.S. et al., 2011. Addressing challenges when studying mobile or episodic species: hierarchical Bayes estimation of occupancy and use. *Journal of Applied Ecology*, 48, pp.56–66.

Morgenstern, H., 1995. Ecologic studies in epidemiology: concepts, principles, and methods. *Annual Review of Public Health*, 16, pp.61–81.

Mosha, J.F. et al., 2014. Hot spot or not: a comparison of spatial statistical methods to predict prospective malaria infections. *Malaria Journal*, 13(1), pp.53–53.

Mullins, G. et al., 1996. Impacts of intensive dairy production on smallholder farm women in coastal Kenya. *Human Ecology*, 24(2), pp.231–253.

Mulu, A., Maier, M. & Liebert, U.G., 2013. Deworming of intestinal helminths reduces HIV-1 subtype C viremia in chronically co-infected individuals. *International Journal of Infectious Diseases*, 17(10), pp.e897–901.

Muok, E.M.O. et al., 2013. Short Report: Association Between CD4(+) T-Lymphocyte Counts and Fecal Excretion of Schistosoma mansoni Eggs in Patients Coinfected with S. mansoni and Human Immunodeficiency Virus Before and After Initiation of Antiretroviral Therapy. *American Journal of Tropical Medicine and Hygiene*, 89(1), pp.42–45.

Mupfasoni, D. et al., 2009. Polyparasite Helminth Infections and Their Association to Anaemia and Undernutrition in Northern Rwanda *PLoS Neglected Tropical Diseases*, 3(9), p.e517.

Murphy, S.P. & Allen, L.H., 2003. Nutritional importance of animal source foods. *Journal of Nutrition*, 133(11 Suppl 2), pp.3932S–3935S.

Murray, C.J.L. et al., 2012. Global malaria mortality between 1980 and 2010: a systematic analysis. *Lancet*, 379(9814), pp.413–431.

Musenge, E. et al., 2013. The contribution of spatial analysis to understanding HIV/TB mortality in children: a structural equation modelling approach. *Global health Action*, 6, p.19266.

Mushayabasa, S. & Bhunu, C.P., 2011. Modeling schistosomiasis and HIV/AIDS codynamics. *Computational and Mathematical Methods in Medicine*, 2011, p.846174.

Mwanakasale, V. et al., 2003. Interactions between Schistosoma haematobium and human immunodeficiency virus type 1: The effects of coinfection on treatment outcomes in rural Zambia. *American Journal of Tropical Medicine and Hygiene*, 69(4), pp.420–428.

Mwandawiro, C.S. et al., 2013. Monitoring and evaluating the impact of national school-based deworming in Kenya: study design and baseline results. *Parasites & Vectors*, 6, p.198.

Mwangi, T.W., Bethony, J.M. & Brooker, S., 2006. Malaria and helminth interactions in humans: an epidemiological viewpoint. *Annals of Tropical Medicine and Parasitology*, 100(7), pp.551–570.

Mwape, K.E. et al., 2012. Taenia solium Infections in a rural area of Eastern Zambia-a community based study. *PLoS Neglected Tropical Diseases*, 6(3), p.e1594.

Mwinzi, P.N.M. et al., 2012. Integrated community-directed intervention for schistosomiasis and soil transmitted helminths in western Kenya - a pilot study. *Parasites & Vectors*, 5, p.182.

Nacher, M., 2005. Comment on: Severe malaria attack is associated with high prevalence of Ascaris

lumbricoides infection among children in rural Senegal. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 99(2), pp.161–3.

Nacher, M. et al., 2000. Ascaris lumbricoides infection is associated with protection from cerebral malaria. *Parasite Immunology*, 22(3), pp.107–113.

Naing C, Whittaker MA, Nyunt-Wai V, Reid SA, Wong SF, Mak JW, Tanner M, 2013. Malaria and soil-transmitted intestinal helminth co-infection and its effect on anaemia: a meta-analysis. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 107, pp. 672–683.

Namagembe, A. et al., 2012. Improved clinical and laboratory skills after team-based, malaria case management training of health care professionals in Uganda. *Malaria Journal*, 11.

Ndeffo Mbah, M.L. et al., 2013. HIV and Schistosoma haematobium prevalences correlate in sub-Saharan Africa. *Tropical Medicine and International Health*, 18(10), pp.1174–1179.

Ndenecho, L., Ndamukong, K.J.N. & Matute, M.M., 2002. Soil transmitted nematodes in children in Buea Health District of Cameroon. *East African Medical Journal*, 79(8), pp.442–445.

Nesbitt, R.A. et al., 2004. Amebiasis and comparison of microscopy to ELISA technique in detection of Entamoeba histolytica and Entamoeba dispar. *Journal of the National Medical Association*, 96(5), pp.671–677.

Neumann, C.G. et al., 2007. Meat supplementation improves growth, cognitive, and behavioral outcomes in Kenyan children. *Journal of Nutrition*, 137(4), pp.1119–1123.

Ngonghala, C.N. et al., 2014. Poverty, disease, and the ecology of complex systems. *PLoS Biology*, 12(4), p.e1001827.

Nguhiu, P.N. et al., 2009. Intestinal polyparasitism in a rural Kenyan community. *East African Medical Journal*, 86(6), pp.272–278.

Nichols, J.D. et al., 2008. Multi-scale occupancy estimation and modelling using multiple detection methods. *Journal of Applied Ecology*, 45, pp.1321–1329.

Nielsen, N.O. et al., 2007. Co-infection with subclinical HIV and Wuchereria bancrofti, and the role of malaria and hookworms, in adult Tanzanians: infection intensities, CD4/CD8 counts and cytokine responses. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 101(6), pp.602–612.

Nielsen, N.O. et al., 2006. Cross-sectional relationship between HIV, lymphatic filariasis and other parasitic infections in adults in coastal northeastern Tanzania. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 100(6), pp.543–550.

Nikolay, B., Brooker, S.J. & Pullan, R.L., 2014. Sensitivity of diagnostic tests for human soil-transmitted helminth infections: a meta-analysis in the absence of a true gold standard. *International Journal for Parasitology*, 44(11), 765-774

Njau, J.D. et al., 2006. Fever treatment and household wealth: the challenge posed for rolling out combination therapy for malaria. *Tropical Medicine and International Health*, 11(3), pp.299–313.

Noblick, J., Skolnik, R. & Hotez, P.J., 2011. Linking global HIV/AIDS treatments with national programs for the control and elimination of the neglected tropical diseases. *PLoS Neglected*

Tropical Diseases, 5(7), p.e1022.

Noor, A.M. et al., 2009. A spatial national health facility database for public health sector planning in Kenya in 2008. *International Journal of Health Geographics*, 8, p.13.

Nundy, S. et al., 2011. Wealth and its associations with enteric parasitic infections in a low-income community in Peru: use of principal component analysis. *American Journal of Tropical Medicine and Hygiene*, 84(1), pp.38–42.

O'Hara, R.B. & Sillanpää, M.J., 2009. A review of Bayesian variable selection methods: what, how and which. *Bayesian Analysis*, 4(1), pp.85–117.

O'Rourke, N., Hatcher, L. & Stepanski, E.J., 2005. *A Step-by-Step Approach to Using SAS for Univariate & Multivariate Statistics*, SAS Institute.

Ochola, L.B. et al., 2006. The reliability of diagnostic techniques in the diagnosis and management of malaria in the absence of a gold standard. *The Lancet Infectious Diseases*, 6(9), pp.582–588.

Odone, A. et al., 2013. Association between Socioeconomic Position and Tuberculosis in a Large Population-Based Study in Rural Malawi. *PLoS ONE*, 8(10), p.e77740.

Ohiokpehai, O. et al., 2007. Feeding patterns and practices among households with children aged 6-59 months in Mbita Division, Suba District, Kenya. *Journal of Food, Agriculture and Environment*, 5, pp.17–23.

Ohrt, C. et al., 2007. Establishing a malaria diagnostics centre of excellence in Kisumu, Kenya. *Malaria Journal*, 6, pp.–79.

Okell, C., 2011. *An analysis of the dynamics of livestock and asset ownership with human health in a rural population in West Kenya.* Unpublished MSc Thesis.

Okell, L.C. et al., 2009. Submicroscopic infection in Plasmodium falciparum-endemic populations: a systematic review and meta-analysis. *Journal of Infectious Diseases*, 200(10), pp.1509–1517.

Okwi, P.O. et al., 2007. Spatial determinants of poverty in rural Kenya. *Proceedings of the National Academy of Sciences of the United States of America*, 104(43), pp.16769–16774.

Olsen, A., Samuelsen, H. & Onyango-Ouma, W., 2001. A study of risk factors for intestinal helminth infections using epidemiological and anthropological approaches. *Journal of Biosocial Science*, 33(4), pp.569–584.

Onda, K., LoBuglio, J. & Bartram, J., 2012. Global access to safe water: accounting for water quality and the resulting impact on MDG progress. *International Journal of Environmental Research and Public Health*, 9(3), pp.880–894.

ONS, 2014. *Mortality Statistics: Deaths Registered in England and Wales (Series DR), 2012. Available at: http://www.ons.gov.uk/ons/rel/vsob1/mortality-statistics--deaths-registered-in-england-and-wales--series-dr-/2012/index.html.*

Open Kenya, 2014. *Kenya Open Data Survey 2014. Available at: https://opendata.go.ke/*

Osorio, F. & Vallejos, R., 2014. SpatialPack: Package for analysis of spatial data. R package version 0.2-3. *cran.r-project.org*. Available at: http://cran.r-project.org/package=SpatialPack.

Otto, G.F., Cort, W. & Keller, A.E., 1931. Environmental studies of families in Tennessee infested with Ascaris, Trichuris and hookworm. *American Journal of Epidemiology*, 14, pp.156–193.

Oxlade, O. & Murray, M., 2012. Tuberculosis and poverty: why are the poor at greater risk in India? *PLoS ONE*, 7(11), p.e47533.

Pacifici, K. et al., 2014. Guidelines for a priori grouping of species in hierarchical community models. *Ecology and Evolution*, 4(7), pp.877–888.

Pai, M., Riley, L.W. & Colford, J.M., 2004. Interferon-gamma assays in the immunodiagnosis of tuberculosis: a systematic review. *The Lancet Infectious Diseases*, 4(12), pp.761–776.

Pai, M., Zwerling, A. & Menzies, D., 2008. Systematic review: T-cell-based assays for the diagnosis of latent tuberculosis infection: An update. *Annals of Internal Medicine*, 149(3), pp.177–184.

Pantaleo, G., Graziosi, C. & Fauci, A.S., 1993. New concepts in the immunopathogenesis of human immunodeficiency virus infection. *New England Journal of Medicine*, 328(5), pp.327–335.

Paradis, E., Claude, J. & Strimmer, K., 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, 20(2), pp.289–290.

Parker, M. & Allen, T., 2011. Does mass drug administration for the integrated treatment of neglected tropical diseases really work? Assessing evidence for the control of schistosomiasis and soil-transmitted helminths in Uganda. *Health Research Policy and Systems*, 9(1), pp.3–3.

Pavlovsky, E.N., 1966. *Natural nidality of transmissible diseases*, University of Illinois Press.

Pawlowski, Z., 2002. Taenia solium: Basic biology and transmission. In G. Singh & S. Prabhakar, eds. *Taenia solium cysticercosis: from basic to clinical science.* CABI International, pp. 1–14.

Payne, D., 1988. Use and limitations of light microscopy for diagnosing malaria at the primary health care level. *Bulletin of the World Health Organization*, 66(5), pp.621–626.

Pearce, N., 1996. Traditional epidemiology, modern epidemiology, and public health. *American Journal of Public Health*, 86(5), pp.678–683.

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, 30(7), pp.683–691.

Pedersen, A.B. & Fenton, A., 2007. Emphasizing the ecology in parasite community ecology. *Trends in Ecology & Evolution*, 22(3), pp.133–139.

Perera, M. et al., 2007. Neglected patients with a neglected disease? A qualitative study of lymphatic filariasis. *PLoS Neglected Tropical Diseases*, 1(2), p.e128.

Peterson, A. T. (2006). Ecologic Niche Modeling and Spatial Patterns of Disease Transmission. *Emerging Infectious Diseases,* 12(12), pp. 1822–1826.

Peters, D.H. & Phillips, T., 2004. Mectizan Donation Program: evaluation of a public-private partnership. *Tropical Medicine and International Health*, 9(4), pp.A4–15.

Petney, T.N. & Andrews, R.H., 1998. Multiparasite communities in animals and humans: frequency, structure and pathogenic significance. *International Journal for Parasitology*, 28(3), pp.377–393.

Petronis, K.R. & Anthony, J.C., 2003. A different kind of contextual effect: geographical clustering of cocaine incidence in the USA. *Journal of Epidemiology and Community Health*, 57(11), pp.893–900.

Pfeiffer, D.U. et al., 2008. *Spatial Analysis in Epidemiology*, OUP Oxford.

Phiri, K. et al., 2000. Urban/rural differences in prevalence and risk factors for intestinal helminth

infection in southern Malawi. *Annals of Tropical Medicine and Parasitology*, 94(4), pp.381–387.

Phongluxa, K. et al., 2014. Perceived illness drives participation in mass deworming campaigns in Laos. *Acta Tropic*a, 141(B)

Pickett, K.E. & Pearl, M., 2001. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *Journal of Epidemiology and Community Health*, 55(2), pp.111–122.

Ponce de Leon, D. et al., 2008. Comparison of an interferon-gamma assay with tuberculin skin testing for detection of tuberculosis (TB) infection in patients with rheumatoid arthritis in a TB-endemic population. *Journal of Rheumatology*, 35(5), pp.776–781.

Popkin, B.M., 1982. A household framework for examining the social and economic consequences of tropical diseases. *Social Science & Medicine*, 16(5), pp.533–543.

Popy, S., Bordignon, L. & Prodon, R., 2009. A weak upward elevational shift in the distributions of breeding birds in the Italian Alps. *Journal of Biogeography*, 37(1), pp.57–67.

Poulin, R., 2004. Macroecological patterns of species richness in parasite assemblages. *Basic and Applied Ecology*, 5(5), pp.423–434.

Poulin, R., 1996. Sexual Inequalities in Helminth Infections: A Cost of Being a Male? *American Naturalist*, 147(2).

Praet, N. et al., 2013. Bayesian modelling to estimate the test characteristics of coprology, coproantigen ELISA and a novel real-time PCR for the diagnosis of taeniasis. *Tropical Medicine and International Health*, 18(5), pp.608–614.

Pullan, R. & Brooker, S., 2008. The health impact of polyparasitism in humans: are we under-estimating the burden of parasitic diseases? *Parasitology*, 135(7), pp.783–794.

Pullan, R.L. et al., 2008. Human Helminth Co-Infection: Analysis of Spatial Patterns and Risk Factors in a Brazilian Community. *PLoS Neglected Tropical Diseases*, 2(12), p.e352.

Pullan, R.L. et al., 2012. Spatial parasite ecology and epidemiology: a review of methods and applications. *Parasitology*, 139(14), pp.1870–1887.

Pullan, R.L., Bukirwa, H., et al., 2010a. Plasmodium infection and its risk factors in eastern Uganda. *Malaria Journal*, 9(2).

Pullan, R.L., Freeman, Matthew C, et al., 2014a. Geographical inequalities in use of improved drinking water supply and sanitation across Sub-Saharan Africa: mapping and spatial analysis of cross-sectional survey data. *PLoS Medicine*, 11(4), p.e1001626.

Pullan, R.L., Gething, P.W., et al., 2011a. Spatial modelling of soil-transmitted helminth infections in Kenya: a disease control planning tool. *PLoS Neglected Tropical Diseases*, 5(2), p.e958.

Pullan, R.L., Kabatereine, N.B., et al., 2011b. Heterogeneities and consequences of Plasmodium species and hookworm coinfection: a population based study in Uganda. *Journal of Infectious Diseases*, 203(3), pp.406–417.

Pullan, R.L., Kabatereine, N.B., et al., 2010b. Spatial and genetic epidemiology of hookworm in a rural community in Uganda. *PLoS Neglected Tropical Diseases*, 4(6), p.e713.

Pullan, R.L., Smith, J.L., et al., 2014b. Global numbers of infection and disease burden of soil

transmitted helminth infections in 2010. *Parasites & Vectors*, 7, p.37.

Quinn, T.C. et al., 2000. Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. *New England Journal of Medicine*, 342(13), pp.921–929.

Quinnell, R.J., 2003. Genetics of susceptibility to human helminth infection. *International Journal for Parasitology*, 33(11), pp.1219–1231.

Quinnell, R.J. et al., 2004. Immune responses in human necatoriasis: association between interleukin-5 responses and resistance to reinfection. *Journal of Infectious Diseases*, 190(3), pp.430–438.

Quinnell, R.J. et al., 1993. Reinfection with hookworm after chemotherapy in Papua New Guinea. *Parasitology*, 106, pp.379–385.

Rajagopal, S., Hotez, P.J. & Bundy, D.A.P., 2014. Micronutrient supplementation and deworming in children with geohelminth infections. *PLoS Neglected Tropical Diseases*, 8(8), p.e2920.

Ramzy, R.M.R. et al., 2006. Effect of yearly mass drug administration with diethylcarbamazine and albendazole on bancroftian filariasis in Egypt: a comprehensive assessment. *Lancet*, 367(9515), pp.992–999.

Randolph, T.F. et al., 2007. Role of livestock in human nutrition and health for poverty reduction in developing countries. *Journal of Animal Science*, 85(11), pp.2788–2800.

Range, N. et al., 2007. HIV and parasitic co-infections in tuberculosis patients: a cross-sectional study in Mwanza, Tanzania. *Annals of Tropical Medicine and Parasitology*, 101(4), pp.343–351.

Raso, G. et al., 2005. Disparities in parasitic infections, perceived ill health and access to health care among poorer and less poor schoolchildren of rural Côte d'Ivoire. *Tropical Medicine and International Health*, 10(1), pp.42–57.

Raso, G. et al., 2004. Multiple parasite infections and their relationship to self-reported morbidity in a community of rural Côte d'Ivoire. *International Journal of Epidemiology*, 33(5), pp.1092–1102.

Raso, G., Vounatsou, P., Gosoniu, L., et al., 2006a. Risk factors and spatial patterns of hookworm infection among schoolchildren in a rural area of western Côte d'Ivoire. *International Journal for Parasitology*, 36(2), pp.201–210.

Raso, G., Vounatsou, P., Singer, B.H., et al., 2006b. An integrated approach for risk profiling and spatial prediction of Schistosoma mansoni-hookworm coinfection. *Proceedings of the National Academy of Sciences of the United States of America*, 103(18), pp.6934–6939.

Redlinger, T. et al., 2002. Hyperendemic Cryptosporidium and Giardia in households lacking municipal sewer and water on the United States-Mexico border. *American Journal of Tropical Medicine and Hygiene*, 66(6), pp.794–798.

Reidpath, D.D., Allotey, P. & Pokhrel, S., 2011. Social sciences research in neglected tropical diseases 2: A bibliographic analysis. *Health Research Policy and Systems*, 9, p.1.

Resende Co, T. et al., 2007. Intestinal helminth co-infection has a negative impact on both anti-Mycobacterium tuberculosis immunity and clinical response to tuberculosis therapy. *Clinical and Experimental Immunology*, 147(1), pp.45–52.

Rezaeian, M. et al., 2007. Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary. *Journal of Epidemiology and Community Health*, 61(2),

pp.98–102.

Rice, A.L. et al., 2000. Malnutrition as an underlying cause of childhood deaths associated with infectious diseases in developing countries. *Bulletin of the World Health Organization*, 78(10), pp.1207–1221.

Richmond, O.M.W., Hines, J.E. & Beissinger, S.R., 2010. Two-species occupancy models: a new parameterization applied to co-occurrence of secretive rails. *Ecological Applications*, 20(7), pp.2036–2046.

Riess, H. et al., 2013. Hookworm infection and environmental factors in mbeya region, Tanzania: a cross-sectional, population-based study. *PLoS Neglected Tropical Diseases*, 7(9), pp.e2408.

Righetti, A.A. et al., 2012. Interactions and potential implications of Plasmodium falciparum-hookworm coinfection in different age groups in south-central Côte d'Ivoire. *PLoS Neglected Tropical Diseases*, 6(11), p.e1889.

Ringshausen, F.C., Schablon, A. & Nienhaus, A., 2012. Interferon-gamma release assays for the tuberculosis serial testing of health care workers: a systematic review. *Journal of Occupational Medicine and Toxicology*, 7.

Rogers, D.J. & Randolph, S.E., 2000. The global spread of malaria in a future, warmer world. *Science*, 289(5485), pp.1763–1766.

Rogier, C., Commenges, D. & Trape, J.F., 1996. Evidence for an age-dependent pyrogenic threshold of Plasmodium falciparum parasitemia in highly endemic populations. *American Journal of Tropical Medicine and Hygiene*, 54(6), pp.613–619.

Roka, M. et al., 2013. Intestinal parasites in HIV-seropositive patients in the Continental Region of Equatorial Guinea: its relation with socio-demographic, health and immune systems factors. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 107(8), pp.502–510.

Roka, M. et al., 2012. Prevalence of intestinal parasites in HIV-positive patients on the island of Bioko, Equatorial Guinea: Its relation to sanitary conditions and socioeconomic factors. *The Science of the Total Environment*, 432, pp.404–411.

Root, E.D. et al., 2013. The role of socioeconomic status in longitudinal trends of cholera in Matlab, Bangladesh, 1993-2007. *PLoS Neglected Tropical Diseases*, 7(1), p.e1997.

Roper, C. et al., 1996. Detection of very low level Plasmodium falciparum infections using the nested polymerase chain reaction and a reassessment of the epidemiology of unstable malaria in Sudan. *American Journal of Tropical Medicine and Hygiene*, 54(4), pp.325–331.

Rota, C.T. et al., 2011. Does accounting for imperfect detection improve species distribution models? *Ecography*, 34, pp.659–670.

Rothman, K.J. & Greenland, S., 1998. *Modern Epidemiology*, Lippincott Williams & Wilkins.

Rothman, K.J., Adami, H.-O. & Ttichopoulos, D., 1998. Should the mission of epidemiology include the eradication of poverty? *The Lancet*, 352(9130), pp.810–813.

Royle, J.A., 2006. Site occupancy models with heterogeneous detection probabilities. *Biometrics*, 62(1), pp.97–102.

Royle, J.A. & Dorazio, R.M., 2008. *Hierarchical Modeling and Inference in Ecology*, Academic

Press.

Royle, J.A. & Link, W.A., 2006. Generalized site occupancy models allowing for false positive and false negative errors. *Ecology*, 87(4), pp.835–841.

Royle, J.A. & Nichols, J.D., 2003. Estimating abundance from repeated presence-absence data or point counts. *Ecology*, 84, pp.777–790.

Royle, J.A. et al., 2012. Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. *Methods in Ecology and Evolution*, 3(3), pp.545–554.

Rutto, J.J. et al., 2013. Socio-economic and cultural determinants of human african trypanosomiasis at the Kenya - Uganda transboundary. *PLoS Neglected Tropical Diseases*, 7(4), p.e2186.

Rwego, I.B. et al., 2008. High rates of Escherichia coli transmission between livestock and humans in rural Uganda. *Journal of Clinical Microbiology*, 46(10), pp.3187–3191.

Saathoff, E. et al., 2005. Ecologic covariates of hookworm infection and reinfection in rural Kwazulu-natal/south Africa: a geographic information system-based study. *American Journal of Tropical Medicine and Hygiene*, 72(4), pp.384–391.

Saaty, T.L., 1980. *The Analytic Hierarchy Process*, McGraw-Hill, New York.

Sachs, J., 2011. *The End of Poverty*, Penguin UK.

Sahn, D.E. & Stifel, D., 2002. Exploring Alternative Measures of Welfare in the Absence of Expenditure Data. *Review of Income and Wealth*, 49(4), pp.463–489.

Saldiva, S.R.M. et al., 2002. Malnutrition and susceptibility to enteroparasites: reinfection rates after mass chemotherapy. *Paediatric and Perinatal Epidemiology*, 16(2), pp.166–171.

Sang, H.C. et al., 2014. Schistosoma haematobium hotspots in south Nyanza, western Kenya: prevalence, distribution and co-endemicity with Schistosoma mansoni and soil-transmitted helminths. *Parasites & Vectors*, 7(1), p.125.

Sangaré, L.R. et al., 2011. Species-specific treatment effects of helminth/HIV-1 co-infection: a systematic review and meta-analysis. *Parasitology*, 138(12), pp.1546–1558.

Santana-Morales, M.A. et al., 2012. Microscopy and molecular biology for the diagnosis and evaluation of malaria in a hospital in a rural area of Ethiopia. *Malaria Journal*, 11.

Sanyaolu, A.O. et al., 2011. Comparative study of entero-parasitic infections among HIV sero-positive and sero-negative patients in Lagos, Nigeria. *Acta Tropica*, 120(3), pp.268–272.

Sanz, J.C. et al., 2006. Application of four Elisa techniques (two for IgM and two for IgG) for serological diagnosis of an outbreak of Q fever. *Enfermedades Infecciosas Y Microbiologia Clinica*, 24(3), pp.178–181.

Sarojinie Fernando, W.T.P. & Hazelton, M.L., 2014. Generalizing the spatial relative risk function. *Spatial and Spatio-Temporal Epidemiology*, 8, pp.1–10.

Sato, Y. et al., 1995. Efficacy of stool examination for detection of Strongyloides infection. *American Journal of Tropical Medicine and Hygiene*, 53(3), pp.248–250.

Sawers, L. & Stillwaggon, E., 2010. Concurrent sexual partnerships do not explain the HIV epidemics in Africa: a systematic review of the evidence. *Journal of the International AIDS Society*, 13(1), pp.34–34.

Schaer, F. et al., 2013. Strongyloides stercoralis: Global Distribution and Risk Factors. *PLoS Neglected Tropical Diseases*, 7(7).

Schantz, P.M. et al., 1992. Neurocysticercosis in an Orthodox Jewish community in New York City. *New England Journal of Medicine*, 327(10), pp.692–695.

Scharlemann, J.P.W. et al., 2008. Global Data for Ecology and Epidemiology: A Novel Algorithm for Temporal Fourier Processing MODIS Data P. Gething, ed. *PLoS ONE*, 3(1), p.e1408.

Schellenberg, J. et al., 2008. Health and survival of young children in southern Tanzania. *BMC Public Health*, 8, p.194.

Schellenberg, J. et al., 2003. Inequities among the very poor: health care for children in rural southern Tanzania. *The Lancet*, 361(9357), pp.561–566.

Schielzeth, H., 2010. Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, 1(2), pp.103–113.

Schmidlin, T. et al., 2013. Effects of Hygiene and Defecation Behavior on Helminths and Intestinal Protozoa Infections in Taabo, Côte d'Ivoire D. J. Diemert, ed. *PLoS ONE*, 8(6), p.e65722.

Schmidt, B.R. et al., 2013. Site occupancy models in the analysis of environmental DNA presence/absence surveys: a case study of an emerging amphibian pathogen. *Methods in Ecology and Evolution*, 4(7), pp.646–653.

Schur, N. et al., 2011. Modelling the geographical distribution of co-infection risk from single-disease surveys. *Statistics in Medicine*, 30(14), pp.1761–1776.

Schüle, S.A. et al., 2014. Ascaris lumbricoides infection and its relation to environmental factors in the Mbeya region of Tanzania, a cross-sectional, population-based study. *PLoS ONE*, 9(3), p.e92032.

Schwartz, S., 1994. The fallacy of the ecological fallacy: the potential misuse of a concept and the consequences. *American Journal of Public Health*, 84(5), pp.819–824.

Scott, M.E., 2008. Ascaris lumbricoides: a review of its epidemiology and relationship to other infections. *Annales Nestle*, 66(1).

Secor, W.E. et al., 2003. Increased density of human immunodeficiency virus type 1 coreceptors CCR5 and CXCR4 on the surfaces of CD4(+) T cells and monocytes of patients with Schistosoma mansoni infection. *Infection and Immunity*, 71(11), pp.6668–6671.

Sester, M. et al., 2011. Interferon-gamma release assays for the diagnosis of active tuberculosis: a systematic review and meta-analysis. *European Respiratory Journal*, 37(1), pp.100–111.

Shapira-Nahor, O. et al., 1998. Increased susceptibility to HIV-1 infection of peripheral blood mononuclear cells from chronically immune-activated individuals. *AIDS*, 12(13), pp.1731–1733.

Shapiro, A.E. et al., 2005. Epidemiology of helminth infections and their relationship to clinical malaria in southwest Uganda. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 99(1), pp.18–24.

Shy, C.M., 1997. The failure of academic epidemiology: witness for the prosecution. *American Journal of Epidemiology*, 145(6), pp.479–484

Silva, C.V.D. et al., 2005. Intestinal parasitic infections in HIV/AIDS patients: experience at a

teaching hospital in central Brazil. *Scandinavian journal of infectious diseases*, 37(3), pp.211–215.

Simarro, P.P. et al., 2012. Estimating and mapping the population at risk of sleeping sickness. *PLoS Neglected Tropical Diseases*, 6(10), pp.e1859.

Simpson, R.E.H., 1952. Infectiousness of communicable diseases in the household (measles, chickenpox, and mumps). *The Lancet*, 2(6734), pp.549–554.

Singer, B.H. & de Castro, Marcia Calclas, 2007. Bridges to sustainable tropical health. *Proceedings of the National Academy of Sciences of the United States of America*, 104(41), pp.16038–16043.

Singer, B.H. & Ryff, C.D., 2007. Neglected tropical diseases, neglected data sources, and neglected issues. *PLoS Neglected Tropical Diseases*, 1(2), p.e104.

Singh, D.D. et al., 2011. Tb or Not Tb? Difficulties in the Diagnosis of Tuberculosis in Hiv-Negative Immigrants to Germany. *European Journal of Medical Research*, 16(9), pp.381–384.

Slaba, K., Skultety, L. & Toman, R., 2005. Efficiency of various serological techniques for diagnosing Coxiella burnetii infection. *Acta Virologica*, 49(2), pp.123–127.

Smith, J.L. & Brooker, S., 2010. Impact of hookworm infection and deworming on anaemia in non-pregnant populations: a systematic review. *Tropical Medicine and International Health*, 15(7), pp.776–795.

Smits, H. et al., 2003. Immunochromatographic Brucella-specific immunoglobulin M and G lateral flow assays for rapid sero-diagnosis of human brucellosis. *Clinical and Diagnostic Laboratory Immunology*, 10(6), pp.1141–1146.

Snijders, T.A.B. & Bosker, R.J., 2011. *Multilevel Analysis*, SAGE.

Soares Magalhães, R.J., Barnett, A.G. & Clements, A.C.A., 2011a. Geographical analysis of the role of water supply and sanitation in the risk of helminth infections of children in West Africa. *Proceedings of the National Academy of Sciences of the United States of America*, 108(50), pp.20084–20089.

Soares Magalhães, R.J., Biritwum, N.-K., et al., 2011b. Mapping helminth co-infection and co-intensity: geostatistical prediction in ghana. *PLoS Neglected Tropical Diseases*, 5(6), p.e1200.

Sodahlon, Y.K. et al., 2013. A success story: Togo is moving toward becoming the first sub-Saharan African nation to eliminate lymphatic filariasis through mass drug administration and countrywide morbidity alleviation. *PLoS Neglected Tropical Diseases*, 7(4), p.e2080.

Sokhna, C. et al., 2004. Increase of malaria attacks among children presenting concomitant infection by Schistosoma mansoni in Senegal. *Malaria Journal*, 3, p.43.

Somi, M.F. et al., 2007. Is there evidence for dual causation between malaria and socioeconomic status? Findings from rural Tanzania. *American Journal of Tropical Medicine and Hygiene*, 77(6), pp.1020–1027.

Speybroeck, N. et al., 2011. True versus Apparent Malaria Infection Prevalence: The Contribution of a Bayesian Approach. *PLoS ONE*, 6(2), pp.–.

Spiegel, A. et al., 2003. Increased frequency of malaria attacks in subjects co-infected by intestinal worms and Plasmodium falciparum malaria. *Transactions of the Royal Society of Tropical*

Medicine and Hygiene, 97(2), pp.198–199.

Spiegel, J.M., 2011. Looking beyond the lamp post: addressing social determinants of neglected tropical diseases in devising integrated control strategies. In *The Causes and Impacts of Neglected Tropical and Zoonotic Diseases: Opportunities for Integrated Intervention Strategies: Workshop Summary*. National Academies Press, p. 604.

Spiegel, J.M. et al., 2010. Which new approaches to tackling neglected tropical diseases show promise? *PLoS Medicine*, 7(5), p.e1000255.

SRA, 2014. *Glossary of risk analysis terms. www.sra.org/sites/default/files/docs/SRA_Glossary.pdf*, Society of Risk Analysis.

Stamey, J.D., Young, D.M. & Seaman, J.W., 2008. A Bayesian approach to adjust for diagnostic misclassification between two mortality causes in Poisson regression. *Statistics in Medicine*, 27(13), pp.2440–2452.

Stanley, S.L., 2003. Amoebiasis. *The Lancet*, 361(9362), pp.1025–1034.

Stark, D. et al., 2014. Evaluation of the EasyScreen™ Enteric Parasite Detection Kit for the detection of Blastocystis spp., Cryptosporidium spp., Dientamoeba fragilis, Entamoeba complex, and Giardia intestinalis from clinical stool samples. *Diagnostic Microbiology and Infectious Disease*, 78(2), pp.149–152.

Stärk, K.D.C. et al., 2006. Concepts for risk-based surveillance in the field of veterinary medicine and veterinary public health: review of current approaches. *BMC Health Services Research*, 6, pp.20–20.

Steinmann, P. et al., 2010. Multiparasitism a neglected reality on global, regional and local scale. *Advances in Parasitology*, 73, pp.21–50.

Steinmann, P. et al., 2007. Occurrence of Strongyloides stercoralis in Yunnan Province, China, and comparison of diagnostic methods. *PLoS Neglected Tropical Diseases*, 1(1), p.e75.

Steinmann, P. et al., 2006. Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk. *The Lancet Infectious Diseases*, 6(7), pp.411–425.

Stevens, K.B. & Pfeiffer, D.U., 2011. Spatial modelling of disease using data- and knowledge-driven approaches. *Spatial and Spatio-temporal Epidemiology*, 2(3), pp.125–133.

Strunz, E.C. et al., 2014. Water, sanitation, hygiene, and soil-transmitted helminth infection: a systematic review and meta-analysis. *PLoS Medicine*, 11(3), p.e1001620.

Stryhn, H. & Christensen, J., 2014. The analysis - hierarchical models: past, present and future. *Preventive Veterinary Medicine*, 113(3), pp.304–312.

Stryhn, H. et al., 2006. Interpretation of variance parameters in multilevel Poisson regression models. In Proceedings of the 11th International Symposium on Veterinary Epidemiology and Economics.

Sturtz, S., Ligges, U. & Gelman, A., 2005. R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software*, 12(3), pp.1–16.

Su, C., Gardner, I. & Johnson, W., 2004. Diagnostic test accuracy and prevalence inferences based on joint and sequential testing with finite population sampling. *Statistics in Medicine*, 23(14),

pp.2237–2255.

Suputtamongkol, Y. et al., 2006. A review of stool ova and parasite examination in the tropics. *Clinical Infectious Diseases*, 43(6), pp.793–794.

Susser, M., 1998. Does risk factor epidemiology put epidemiology at risk? Peering into the future. *Journal of Epidemiology and Community Health*, 52(10), pp.608–611.

Susser, M., 1994. The logic in ecological: I. The logic of analysis. *American Journal of Public Health*, 84(5), pp.825–829.

Susser, M. & Susser, E., 1996. Choosing a future for epidemiology: II. From black box to Chinese boxes and eco-epidemiology. *American Journal of Public Health*, 86(5), pp.674–677.

Swaminathan, S. et al., 2012. Epidemiological Assessment of Eight Rounds of Mass Drug Administration for Lymphatic Filariasis in India: Implications for Monitoring and Evaluation. *PLoS Neglected Tropical Diseases*, 6(11).

Tanadini, L.G. & Schmidt, B.R., 2011. Population size influences amphibian detection probability: implications for biodiversity monitoring programs. *PLoS ONE*, 6(12), p.e28244.

Tanser, F. et al., 2009. Localized spatial clustering of HIV infections in a widely disseminated rural South African epidemic. *International Journal of Epidemiology*, 38(4), pp.1008–1016.

Tarafder, M.R. et al., 2010. Estimating the sensitivity and specificity of Kato-Katz stool examination technique for detection of hookworms, Ascaris lumbricoides and Trichuris trichiura infections in humans in the absence of a 'gold standard'. *International Journal for Parasitology*, 40(4), pp.399–404.

Taylor-Robinson, D.C. et al., 2012. Deworming drugs for soil-transmitted intestinal worms in children: effects on nutritional indicators, haemoglobin and school performance. *Cochrane Database of Systematic Reviews*, 11, p.CD000371.

Tchuem Tchuenté, L.-A. et al., 2003. Polyparasitism with Schistosoma haematobium and soil-transmitted helminth infections among school children in Loum, Cameroon. *Tropical Medicine and International Health*, 8(11), pp.975–986.

Thomas, L.F., 2013. *Epidemiology of Taenia solium cysticercosis in western Kenya*. Unpublished PhD thesis

Thomas, T.A. et al., 2010. Malnutrition and Helminth Infection Affect Performance of an Interferon {gamma}-Release Assay. *Pediatrics*, 126(6), p.e1522.

Thurnham, D.I. et al., 2010. Adjusting plasma ferritin concentrations to remove the effects of subclinical inflammation in the assessment of iron deficiency: a meta-analysis. *American Journal of Clinical Nutrition*, 92(3), pp.546–555.

Thurnham, D.I. et al., 2003. Effects of subclinical infection on plasma retinol concentrations and assessment of prevalence of vitamin A deficiency: meta-analysis. *Lancet*, 362(9401), pp.2052–2058.

Torres-Gonzalez, P. et al., 2013. Prevalence of Latent and Active Tuberculosis among Dairy Farm Workers Exposed to Cattle Infected by Mycobacterium bovis. *PLoS Neglected Tropical Diseases*, 7(4).

Traore, M.O. et al., 2012. Proof-of-principle of onchocerciasis elimination with ivermectin treatment in endemic foci in Africa: final results of a study in Mali and Senegal. *PLoS Neglected Tropical Diseases*, 6(9), p.e1825.

Trape, J.F., 1985. Rapid Evaluation of Malaria Parasite Density and Standardization of Thick Smear Examination for Epidemiological Investigations. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 79(2), pp.181–184.

Tristão-Sá, R. et al., 2002. Intestinal nematodes and pulmonary tuberculosis. *Revista da Sociedade Brasileira de Medicina Tropical*, 35(5), pp.533–535.

Trönnberg, L. et al., 2010. Household-based prevalence of helminths and parasitic protozoa in rural KwaZulu-Natal, South Africa, assessed from faecal vault sampling. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 104(10), pp.646–652.

Tshikuka, J.G. et al., 1996. Multiple infection with Plasmodium and helminths in communities of low and relatively high socio-economic status. *Annals of Tropical Medicine and Parasitology*, 90(3), pp.277–293.

Tyre, A.J. et al., 2003. Improving precision and reducing bias in biological surveys: Estimating false-negative error rates. *Ecological Applications,* 13(6), pp.1790–1801.

Ukwuani, F.A. & Suchindran, C.M., 2003. Implications of women's work for child nutritional status in sub-Saharan Africa: a case study of Nigeria. *Social Science & Medicine*, 56(10), pp.2109–2121.

UNAIDS, 2014. *Global AIDS Response Reporting, 2014: http://www.unaids.org/en/dataanalysis/knowyourresponse/globalaidsprogressreporting*, Geneva: Joint United Nations Programme on HIV/AIDS.

UNDP, 2013. *MDG report 2013: Assessing progress in Africa towards the Millenium Development Goals. United Nations, Geneva*,

Uparanukraw, P., Phongsri, S. & Morakote, N., 1999. Fluctuations of larval excretion in Strongyloides stercoralis infection. *American Journal of Tropical Medicine and Hygiene*, 60(6), pp.967–973.

UtCNTD, 2014. *Delivering on Promises and Driving Progress - Second Progress Report.* Available at: *http://unitingtocombatntds.org/resource/delivering-promises-and-driving-progress-second-progress-report*.

Utzinger, J. et al., 2008. FLOTAC: a new sensitive technique for the diagnosis of hookworm infections in humans. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 102(1), pp.84–90.

Utzinger, J. et al., 2011. From innovation to application: social-ecological context, diagnostics, drugs and integrated control of schistosomiasis. *Acta Tropica*, 120 Suppl 1, pp.S121–37.

Utzinger, J. et al., 1999. Intestinal amoebiasis, giardiasis and geohelminthiases: their association with other intestinal parasites and reported intestinal symptoms. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 93(2), pp.137–141.

Utzinger, J. et al., 2001. Relative contribution of day-to-day and intra-specimen variation in faecal egg

counts of Schistosoma mansoni before and after treatment with praziquantel. *Parasitology*, 122(Pt 5), pp.537–544.

Utzinger, J. et al., 2009. Schistosomiasis and neglected tropical diseases: towards integrated and sustainable control and a word of caution. *Parasitology*, 136(13), pp.1859–1874.

Utzinger, J. et al., 2003. Sustainable schistosomiasis control - the way forward. *The Lancet*, 362(9399), pp.1932–1934.

van den Bogaart, E. et al., 2012. Prevalence, features and risk factors for malaria co-infections amongst visceral leishmaniasis patients from Amudat Hospital, Uganda. *PLoS Neglected Tropical Diseases*, 6(4), p.e1617.

van Geertruyden, J.P. & D'Alessandro, U., 2007. Malaria and HIV: a silent alliance. *Trends in Parasitology*, 23(10), pp.465–467.

van Oosterhout A.J.M., Motta A.C., 2005. Th1/Th2 paradigm: not seeing the forest for the trees? *European Respiratory Journal*, 25, pp.591–593

van Wyk, J.A., 2001. Refugia- overlooked as perhaps the most potent factor concerning the development of anthelmintic resistance. *Onderstepoort Journal of Veterinary Research*, 68(1), pp.55–67.

Vanderelst, D. & Speybroeck, N., 2010. Quantifying the lack of scientific interest in neglected tropical diseases. *PLoS Neglected Tropical Diseases*, 4(1), p.e576.

Vercruysse, J. et al., 2011. Is anthelmintic resistance a concern for the control of human soil-transmitted helminths? *International Journal for Parasitology,* 1(1), pp.14–27.

Vercruysse, J., Levecke, B. & Prichard, R., 2012. Human soil-transmitted helminths: implications of mass drug administration. *Current Opinion in Infectious Diseases*, 25(6), pp.703–708.

Vergne, T. et al., 2014. Zero-inflated models for identifying disease risk factors when case detection is imperfect: Application to highly pathogenic avian influenza H5N1 in Thailand. *Preventive Veterinary Medicine*, 114(1), pp.28–36.

Victora, C.G. et al., 1997. The role of conceptual frameworks in epidemiological analysis: a hierarchical approach. *International journal of epidemiology*, 26(1), pp.224–227.

Victora, C.G.C. et al., 1994. Risk factors for pneumonia among children in a Brazilian metropolitan area. *Pediatrics*, 93(6), pp.977–985.

Vijayakumar, T.-S. et al., 2005. Performance of a rapid immunochromatographic screening test for detection of antibodies to human immunodeficiency virus type 1 (HIV-1) and HIV-2: experience at a tertiary care hospital in South India. *Journal of Clinical Microbiology*, 43(8), pp.4194–4196.

Vilhena, M., Santos, M. & Torgal, J., 1999. Seroprevalence of human cysticercosis in Maputo, Mozambique. *American Journal of Tropical Medicine and Hygiene*, 61(1), pp.59–62.

Viney, M.E. et al., 2004. Why does HIV infection not lead to disseminated strongyloidiasis? *Journal of Infectious Diseases*, 190(12), pp.2175–2180.

Vounatsou, P. et al., 2009. Bayesian geostatistical modelling for mapping schistosomiasis transmission. *Parasitology*, 136(13), pp.1695–1705.

Vyas, S. & Kumaranayake, L., 2006. Constructing socio-economic status indices: how to use

principal components analysis. *Health Policy and Planning*, 21(6), pp.459–468.

Wahyuni, S. et al., 2004. Clustering of filarial infection in an age-graded study: genetic, household and environmental influences. *Parasitology*, 128(Pt 3), pp.315–321.

Waikagul, J. et al., 2002. A cross-sectional study of intestinal parasitic infections among schoolchildren in Nan Province, Northern Thailand. *Southeast Asian Journal of Tropical Medicine and Public Health*, 33(2), pp.218–223.

Waithaka, M. et al., 2002. *Characterization of dairy systems in the western Kenya region*, Smallholder Dairy (R&D) Project.

Walker, M., Hall, A. & Basanez, M.-G., 2011. Individual Predisposition, Household Clustering and Risk Factors for Human Infection with Ascaris lumbricoides: New Epidemiological Insights *PLoS Neglected Tropical Diseases*, 5(4), p.e1047.

Walson, J.L. & John-Stewart, G., 2007. Treatment of helminth co-infection in individuals with HIV-1: A systematic review of the literature. *PLoS Neglected Tropical Diseases*, 1(3), p.e102.

Walson, J.L., Herrin, B.R. & John-Stewart, G., 2009. Deworming helminth co-infected individuals for delaying HIV disease progression. *Cochrane Database of Systematic Reviews*, (3), p.CD006419.

Wamae, N. et al., 2006. Community-directed treatment of lymphatic filariasis in Kenya and its role in the national programmes for elimination of lymphatic filariasis. *African Journal of Health Sciences*, 13(1-2), pp.69–79.

Wang, X., 2012. Soil-transmitted helminth infections and correlated risk factors in preschool and school-aged children in rural Southwest China. *PLoS One*, 7(9), e45939.

Warren, K.S., 1981. The control of helminths: nonreplicating infectious agents of man. *Annual Review of Public Health*, 2, pp.101–105.

Webb, E.L. et al., 2011. Effect of single-dose anthelmintic treatment during pregnancy on an infant's response to immunisation and on susceptibility to infectious diseases in infancy: a randomised, double-blind, placebo-controlled trial. *Lancet*, 377(9759), pp.52–62.

Webb, E.L., Ekii, A.O. & Pala, P., 2012. Epidemiology and immunology of helminth-HIV interactions. *Current opinion in HIV and AIDS*, 7(3), pp.245–253.

Wenger, S.J. & Freeman, Mary C, 2008. Estimating species occurrence, abundance, and detection probability using zero-inflated distributions. *Ecology*, 89(10), pp.2953–2959.

West, P.A. et al., 2013. Malaria risk factors in North West Tanzania: the effect of spraying, nets and wealth. *PLoS ONE*, 8(6), p.e65787.

Whitehead, A., 2006. Persistent poverty in North East Ghana. *Journal of Development Studies*, 42(2), pp.278–300.

Whitehead, M., Dahlgren, G. & Evans, T., 2001. Equity and health sector reforms: can low-income countries escape the medical poverty trap? *The Lancet*, 358(9284), pp.833–836.

WHO, 2012a. *Accelerating work to overcome the global impact of neglected tropical diseases*, Geneva : World Health Organization.

WHO, 2013a. *Control and surveillance of human African trypanosomiasis: Report of WHO Expert Committee*, Geneva : World Health Organization.

WHO, 2012b. *Global Report for Research on Infectious Diseases of Poverty*, Geneva : World Health Organization.

WHO, 2013b. *Global Tuberculosis Report 2013.* Available at: *http://apps.who.int/iris/handle/10665/91355*, Geneva: World Health Organization.

WHO, 2011. *Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity. Vitamin and Mineral Nutrition Information System*, Geneva: WHO.

WHO, 2002. *Independent participatory monitoring of CDTI projects. Guidelines and instruments. African Programme for Onchocerciasis Control* http://www.who.int/entity/apoc/publications/cdti_monitoring/en/, Geneva: WHO.

WHO, 2012c. *London declaration on neglected tropical diseases: uniting to combat NTDs. www.who.int/neglected_diseases/London_Declaration_NTDs.pdf*,

WHO, 2006a. *Meeting the MDG drinking water and sanitation target: the urban and rural challenge of the decade*, Geneva : World Health Organization.

WHO, 2006b. *Preventive chemotherapy in human helminthiasis. Coordinated use of anthelminthic drugs in control interventions: A manual for health professionals and programme managers*, Geneva: WHO.

WHO, 2010. *Working to overcome the global impact of neglected tropical diseases. First WHO report on neglected tropical diseases*, Geneva: WHO.

Williams-Blangero, S. et al., 1999. Genetic analysis of susceptibility to infection with Ascaris lumbricoides. *American Journal of Tropical Medicine and Hygiene*, 60(6), pp.921–926.

Wilson, D. & Halperin, D.T., 2008. "Know your epidemic, know your response": a useful approach, if we get it right. *The Lancet*, 372(9637), pp.423–426.

Wiria, A.E. et al., 2013. The effect of three-monthly albendazole treatment on malarial parasitemia and allergy: a household-based cluster-randomized, double-blind, placebo-controlled trial. *PLoS ONE*, 8(3), p.e57899.

Wolday, D. et al., 2002. Treatment of intestinal worms is associated with decreased HIV plasma viral load. *Journal of Acquired Immune Deficiency Syndromes*, 31(1), pp.56–62.

Wolstenholme, A.J. et al., 2004. Drug resistance in veterinary helminths. *Trends in Parasitology*, 20(10), pp.469–476.

Wongsrichanalai, C. et al., 2007. A review of malaria diagnostic tools: Microscopy and rapid diagnostic test (RDT). *American Journal of Tropical Medicine and Hygiene*, 77(6), pp.119–127.

Woo, P.T., 1971. Evaluation of the haematocrit centrifuge and other techniques for the field diagnosis of human trypanosomiasis and filariasis. *Acta Tropica*, 28(3), pp.298–303.

Woolhouse, M.E. et al., 1997. Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proceedings of the National Academy of Sciences of the United States of America*, 94(1), pp.338–342.

World Bank, 2014. *http://povertydata.worldbank.org/poverty/region/SSA*.

World Bank, 2012. *Kenya - Additional Financing for the Water and Sanitation Services Improvement Project*. Available at: http://www.worldbank.org/projects/P126637/water-sanitation-service-

improvement-project-additional-financing?lang=en.

Worrall, E., Basu, S. & Hanson, K., 2005. Is malaria a disease of poverty? A review of the literature. *Tropical Medicine and International Health*, 10(10), pp.1047–1059.

Wouters, O.J. et al., 2014. Resource planning for neglected tropical disease (NTD) control programs: feasibility study of the Tool for Integrated Planning and Costing (TIPAC). *PLoS Neglected Tropical Diseases*, 8(2), pp.e2619–e2619.

Woyessa, A. et al., 2013. Malaria risk factors in Butajira area, south-central Ethiopia: a multilevel analysis. *Malaria Journal*, 12, p.273.

Wyss, K., Hutton, G. & N'Diekhor, Y., 2004. Costs attributable to AIDS at household level in Chad. *AIDS care*, 16(7), pp.808–816.

Xu, K. et al., 2003. Household catastrophic health expenditure: a multicountry analysis. *The Lancet*, 362(9378), pp.111–117.

Xue, Z. et al., 2011. Impact of temperature and precipitation on propagation of intestinal schistosomiasis in an irrigated region in Ethiopia: suitability of satellite datasets. *Tropical Medicine and International Health*, 16(9), pp.1104–1111.

Yap, P. et al., 2013. Rapid re-infection with soil-transmitted helminths after triple-dose albendazole treatment of school-aged children in Yunnan, People's Republic of China. *American Journal of Tropical Medicine and Hygiene*, 89(1), pp.23–31.

Yapi, R.B. et al., 2014. Infection and co-infection with helminths and Plasmodium among school children in Côte d'Ivoire: results from a National Cross-Sectional Survey. *PLoS Neglected Tropical Diseases*, 8(6), p.e2913.

Zaralis, K. et al., 2009. Consequences of protein supplementation for anorexia, expression of immunity and plasma leptin concentrations in parasitized ewes of two breeds. *British Journal of Nutrition*, 101(4), pp.499–509.

Zelner, J.L. et al., 2012. Social connectedness and disease transmission: social organization, cohesion, village context, and infection risk in rural Ecuador. *American Journal of Public Health*, 102(12), pp.2233–2239.

Zere, E. & McIntyre, D., 2003. Equity in self-reported adult illness and use of health service in South Africa: inter-temporal comparison. *Journal of Health Population and Nutrition*, 21(3), pp.205–215.

Zhou, Y.-B. et al., 2014. HIV-, HCV-, and co-infections and associated risk factors among drug users in southwestern China: a township-level ecological study incorporating spatial regression. *PLoS ONE*, 9(3), p.e93157.

Zhou, Y.-B. et al., 2013. Spatial-temporal variations of Schistosoma japonicum distribution after an integrated national control strategy: a cohort in a marshland area of China. *BMC Public Health*, 13.

Ziegelbauer, K. et al., 2012. Effect of Sanitation on Soil-Transmitted Helminth Infection: Systematic Review and Meta-Analysis. *PLoS Medicine*, 9(1).

Zinsstag, J. et al., 2011. From "one medicine" to “one health” and systemic approaches to health and

well-being. *Preventive Veterinary Medicine*, 101(3-4), pp.148–156.

Zipkin, E.F. et al., 2010. Multi-species occurrence models to evaluate the effects of conservation and management actions. *Biological Conservation*, 143(2), pp.479–484.

Zipkin, E.F., DeWan, A. & Royle, J.A., 2009. Impacts of forest fragmentation on species richness: a hierarchical approach to community modelling. *Journal of Applied Ecology*, 46(4), pp.815–822.

Zulu, L.C., Kalipeni, E. & Johannes, E., 2014. Analyzing spatial clustering and the spatiotemporal nature and trends of HIV/AIDS prevalence using GIS: the case of Malawi, 1994-2010. *BMC Infectious Diseases*, 14, p.285.

Zuur, A.F., Saveliev, A.A. & Ieno, E.N., 2012. *Zero Inflated Models and Generalized Linear Mixed Models with R*, Highland Statistics Limited.

Zwerling, A. et al., 2013. Repeat IGRA Testing in Canadian Health Workers: Conversions or Unexplained Variability? *PLoS ONE*, 8(1).